N° D'ORDRE :

403

THÈSE

PRÉSENTÉE A

L'UNIVERSITÉ DE BORDEAUX I

POUR OBTENIR LE GRADE DE

DOCTEUR ÈS SCIENCES PHYSIQUES

PAR

Bernard BUSETTA

METHODES DIRECTES EN RADIOCRISTALLOGRAPHIE
ET RECHERCHE SUR LES HORMONES STEROIDES

Soutenue le 26 mai 1973 devant la Commission d'Examen :

MM. R. GAY,	Professeur	*Président*
EE. BAULIEU,	Professeur à l'Université de Paris XI	*Examinateurs*
A. PACAULT,	Professeur	
J. CLASTRE,	Professeur	
M. HOSPITAL,	Docteur ès Sciences	

— 1973 —

Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41547324_0024

N° D'ORDRE :

403

THÈSE

PRÉSENTÉE A

L'UNIVERSITÉ DE BORDEAUX I

POUR OBTENIR LE GRADE DE

DOCTEUR ÈS SCIENCES PHYSIQUES

PAR

Bernard BUSETTA

METHODES DIRECTES EN RADIOCRISTALLOGRAPHIE
ET RECHERCHE SUR LES HORMONES STEROIDES

Soutenue le 26 mai 1973 devant la Commission d'Examen :

MM. R. GAY, Professeur		*Président*
EE. BAULIEU,	Professeur à l'Université de Paris XI	*Examinateurs*
A. PACAULT,	Professeur	
J. CLASTRE,	Professeur	
M. HOSPITAL,	Docteur ès Sciences	

— 1973 —

UNIV.
LEIDEN
BIBL.

A mes Parents

Ce travail a été réalisé au laboratoire de Cristallographie de l'Université de Bordeaux I.

Nous prions Monsieur le Professeur R.GAY de trouver ici le témoignage de notre profonde et respectueuse gratitude pour sa constante bienveillance à notre égard et pour l'intérêt qu'il n'a cessé de porter à ce travail.

Monsieur le Professeur E.E. BAULIEU a été notre directeur scientifique dans le domaine biologique. Les fréquentes et fructueuses discussions que nous avons eues avec lui, nous ont évité bien des erreurs et ont permis une progression plus rapide de ce travail.

Nous pensons qu'il est inutile de souligner qu'une thèse représente, maintenant, plus le travail d'un groupe que celui d'un chercheur isolé. Au fil des années, s'est établi avec Messieurs M.HOSPITAL et C.COURSEILLE cet esprit d'équipe qui est une des caractéristiques de la recherche moderne. Cette coopération nous a été très précieuse.

Nous remercions également tous ceux qui ont participé, à nos côtés, à cette recherche : Mme F.LEROY, Messieurs G.COMBERTON, J.M.FORNIES MARQUINA, G.PRECIGOUX.

Nous tenons également à remercier très vivement, Monsieur R.WEISS, Professeur à la Faculté des Sciences de Strasbourg, qui a bien voulu accepter d'être notre parrain auprès du Centre National de la Recherche Scientifique.

Monsieur le Professeur A.PACAULT nous a fait l'honneur de participer à ce jury; qu'il trouve ici l'expression de notre respectueuse gratitude.

Nous remercions également Monsieur le Professeur J.CLASTRE, membre de ce jury, et Monsieur J.C.COLLETER Professeur à l'Université de Bordeaux II pour la bienveillante attention qu'ils ont toujours manifesté à mon égard.

A tous les chercheurs du laboratoire qui nous ont entouré de leur sympathie et au personnel technique dont l'aimable et efficace collaboration a permis la réalisation pratique de cet ouvrage, nous témoignons nos sentiments très reconnaissants.

SOMMAIRE

PREMIERE SECTION : A LA RECHERCHE D'UNE METHODE DE DETERMINATION DE STRUCTURES CRISTALLINES DE COMPOSES BIOLOGIQUEMENT ACTIFS

INTRODUCTION

PREMIERE PARTIE : PREMIERE APPROCHE DES METHODES DIRECTES

INTRODUCTION

CHAPITRE I : Généralités sur les méthodes directes

I-1 Position du problème
I-2 Mode Général de Résolution
I-3 Les Formules de la Tangente

CHAPITRE II : Méthode d'addition symbolique des phases

II-1 Stratégie Initiale
II-2 Réalisation de Programmes d'Addition Symbolique pour IBM-1620
II-3 Application aux structures "pseudo centrées" - Recherche d'une stratégie expérimentale

II-3-1 Analyse de la Génération des Phases pour l'oestradiol ($P2_12_12$)
II-3-2 Première Stratégie Expérimentale

II-4 Le Contrôle Symbolique des Phases
II-5 Conclusion

CHAPITRE III : Méthode de la Multisolution

III-1 Stratégie Initiale (MM.Woolfson, G.Germain, P.Main)
III-2 Utilisation Simultanée de la Multisolution et de l'addition symbolique
III-3 Les Figures de Mérite
III-4 Application à la Résolution de Structures Stéroïdiques

CHAPITRE IV : Méthode des Invariants de Structure

IV-1 Principe de la méthode
IV-2 Etude Expérimentale

IV-3 Quelques Réflexions sur le Facteur d'échelle
IV-4 Essais d'Amélioration du calcul des cosinus invariants

IV-4-1 La formule MDKS de H.Hauptman
IV-4-2 Utilisation des relations entre invariants
IV-4-3 Pondération semi empirique

IV-5 Calcul des Phases et Symboles
IV-6 Application à la détermination de l'androsténédione
IV-7 Conclusion

CONCLUSION

DEUXIEME PARTIE : NOUVELLES CONCEPTIONS, NOUVELLES METHODES

INTRODUCTION

CHAPITRE I : Interprétation Physique de la Génération de Phases par Méthodes Directes et Applications

I-1 Invariants de structure
I-2 Choix de l'origine et Génération des Phases
I-3 Influence de la fonction maximisante de la formule de la tangente

CHAPITRE II : Notions d'iles moléculaires

II-1 Existence d'iles moléculaires
II-2 Influence du pouvoir de résolution

CHAPITRE III : Mode de Résolution des Systèmes à iles moléculaires

III-1 Cas où une des molécules est connue

III-1-1 Technique de Résolution Générale
III-1-2 Etude de Structures à atomes lourds
III-1-3 Exemple d'Application pratique

III-2 Etude du cas général

III-2-1 Choix des symboles
III-2-2 Figures de Mérite à priori
III-2-3 Nouvelles Considérations sur les figures de mérite (à postériori)
III-2-4 Applications Pratiques

III-3 Conclusion

CHAPITRE IV : Iles de Groupes

IV-1 Le phénomène d'iles dans le groupe $P\bar{1}$

IV-1-1 Existence d'iles de groupe
IV-1-2 Dépendance du choix de l'origine

IV-1-3 Schéma Général des iles de groupe
IV-1-4 Intersection

IV-2 Problèmes Posés par les Iles de Groupe
IV-3 Exemples de Résolution

CONCLUSION

* * * *

ANNEXE : OBSERVATIONS EXPERIMENTALES

* * * *

DEUXIEME SECTION : ETUDE DE LA RELATION STRUCTURE-ACTIVITE DANS LE CAS DES HORMONES OESTROGENES

PREMIERE PARTIE : OESTRADIOL ET DIETHYLSTILBOESTROL

INTRODUCTION

CHAPITRE I : Les deux conformations du diéthylstilboestrol

I-1 Introduction
I-2 Le diéthylstilboestrol centrosymétrique
I-3 Le diéthylstilboestrol à symétrie pseudobinaire
I-4 Etude de la transition $DES_{\bar{1}} \rightarrow DES_{"2"}$
I-5 Structure du Cycladiène
I-6 Conclusion

CHAPITRE II : Comparaison de la cohésion cristalline dans les structures du DES et de l'oestradiol

II-1 Les structures Orthorhombiques Pbca du DES et du cycladiène

II-1-1 Etude de la cohésion cristalline suivant b, c
II-1-2 Cohésion cristalline suivant a

II-2 Les structures solvatées du DES de Symétrie Triclinique

II-2-1 DES + CH_3OH + H_2O
II-2-2 DES + C_2H_5OH
II-2-3 DES + diméthyl sulfoxyde

II-3 Conclusion

CHAPITRE III : Relation Structure-Activité DES-Oestradiol

III-1 Avant Propos
III-2 Activité Possible du DES $\bar{I}$
III-3 Activité Possible du DES$_{"2"}$
III-4 Le Problème des dérivés du Triphényl éthylène
III-5 Essais "d'amélioration" des molécules d'oestradiol et de DES

CONCLUSION

DEUXIEME PARTIE : QUELQUES REFLEXIONS SUR LES INTERACTIONS MOLECULAIRES DES COMPOSES OESTROGENES ET ANDROGENES

CHAPITRE I : Etudes des Interactions moléculaires au niveau de l'oxygène O_3 du cycle A

I-1 Liaisons hydrogène faibles (Stéroïdes Androgènes)
I-2 Liaisons hydrogène fortes (Stéroïdes Oestrogènes)
I-3 Etude de l'Intéraction DES-ℓ Tyrosine Ethyl Ester
I-4 Quelques objections aux hypothèses précédentes

CHAPITRE II : Etudes des Interactions moléculaires au niveau de la fonction 17 β ol

II-1 Stabilité de l'association stéroïde-eau au niveau de la fonction 17 β ol
II-2 Situation de la molécule d'eau par rapport au stéroïde
II-3 Hypothèses sur le rôle de la fonction 17 β ol
II-4 Conclusion

CONCLUSION

PREMIERE SECTION

A LA RECHERCHE D'UNE METHODE DE DETERMINATION DE STRUCTURES CRISTALLINES DE COMPOSES BIOLOGIQUEMENT ACTIFS

A LA RECHERCHE D'UNE METHODE DE DETERMINATION DE STRUCTURES CRISTALLINES DE COMPOSES BIOLOGIQUEMENT ACTIFS

INTRODUCTION

Actuellement un pharmacologue voulant analyser un corps actif nouveau s'adresse de préférence au chimiste,plutôt qu'au cristallographe. Il faut reconnaitre que,jusqu'ici, ce dernier pouvait, dans une série de corps actifs, déterminer seulement les structures cristallines de composés qu'il savait résoudre et non pas n'importe quelle structure demandée par le pharmacologue; au contraire le chimiste, dans un temps plus ou moins long, pouvait parvenir à déterminer la configuration moléculaire.

Or la cristallographie présente de nombreux avantages sur l'analyse chimique; notamment :

- elle nécessite peu de produit et généralement celui-ci n'est pas détérioré.
- elle donne des résultats beaucoup plus précis sur la géométrie de la molécule.
- elle peut fournir des schémas sur le mode d'interactions du composé dans le milieu biologique.

Mais pour intéresser le pharmacologue, il faut pouvoir déterminer n'importe quelle structure le plus rapidement possible.

Vouloir déterminer des structures de composés biologiquement actifs en série c'est forcément mettre au point de nouvelles méthodes de recherches ou peut être trouver des variantesde méthodes déjà existantes, qui utiliseraient au mieux les possibilités de chacune d'elles.

Le premier objectif de ce travail serait donc de mettre au point un mode de résolution, permettant d'obtenir, pour n'importe quelle structure cristalline,une hypothèse de départ directement affinable par moindres carrés, mais ne nécessitant pas cependant un temps de calcul sur ordinateur trop élevé.

* *

*

Jusqu'ici beaucoup de structures cristallines ont été déterminées par la méthode de l'atome lourd,en utilisant des dérivés bromés, chlorés, ... Cependant de tels groupements sont généralement incompatibles avec la notion de corps actif. La configuration moléculaire du composé véritable apparait fortement perturbée, et l'arrangement cristallin ne peut plus fournir de schéma valable du mode d'interactions de la molécule dans le milieu biologique. Dans la plupart des cas, les composés actifs sont constitués uniquement d'atomes de même "poids" (carbone, azote, oxygène) et nécessitent des techniques de recherches plus élaborées.

Notons de plus,qu'un composé actif agit généralement sur la lumière polarisée : il est lévogyre. Dès lors, il n'existe souvent dans la maille cristalline que le seul énanthiomorphe actif : le groupe de symétrie ne comportant pas d'éléments inverses, on a affaire à une structure cristalline non centrée du type P_1, P_2, P_{222}, P_3...

Or si, à ce jour, les méthodes de résolution des structures (centrosymétriques)à atomes identiques ont fait l'objet d'une étude à peu près complète, (les méthodes directes dans ce domaine sont maintenant pratique courante et nous les supposerons connues), il en est tout autrement pour les structures non centrées. Aucun mode d'investigation n'a,semble-t-il,réellement fait ses preuves, qu'il s'agisse des méthodes indirectes ou des nombreuses méthodes directes proposées à ce jour.

La limitation du temps de calcul que nous nous sommes imposés au départ nous a conduit à éviter les méthodes indirectes basées sur l'analyse automatique de la fonction de Patterson tridimensionnelle, malgré les succès qu'elles semblaient obtenir au moment où

nous avons entrepris ce travail (*) . Le temps de calcul nécessaire est, dans les meilleurs cas, encore trop élevé (en moyenne 30 mn de CDC 6600 c'est à dire plus de 3 h. d'IBM 360-44), et de plus, ces méthodes,nécessitant le stockage en mémoire de l'espace de Patterson, demandent des ordinateurs de grande capacité. En outre,elles réclament une connaissance approximative de la configuration de la molécule ou,tout au moins,d'une partie de celle-ci.

Pour les mêmes raisons,nous n'avons pas pris en considération la méthode de minimisation de l'énergie potentielle dans la maille cristalline, qui avait permis à E.Giglio (24) de réussir la structure du 5α androstane 3-17 dione.

* *
*

Seules les méthodes directes sont réellement adaptées à notre problème : elles ne nécessitent aucune connaissance à priori de la structure, permettent l'étude de molécules non rigides, et surtout de structures cristallines de complexes (corps actifs + éléments protéïniques) qui constituent une étape indispensable pour comprendre la fixation du corps actif sur les proteïnes réceptrices (ou de transport).

Ces méthodes ont cependant à nos yeux un inconvénient: leur fondements théoriques font appel à des notions mathématiques relativement compliquées. Comme elles s'appliquent à des phénomènes physiques,on peut se demander s'il n'existerait pas une explication plus concrète des processus statistiques qu'elles introduisent. L'utilisateur de ces méthodes n'étant pas censé être un mathématicien, nous avons voulu trouver une interprétation physique satisfaisante pouvant servir de support à l'analyse statistique.

Il est difficile de choisir,à priori,pour l'utilisation des méthodes directes, une technique plutôt qu'une autre. Seules, une longue expérience et l'analyse minutieuse des causes d'échecs des techniques habituelles permettront de mettre au point une mode de résolution efficace. C'est ce lent cheminement,où chaque nouvelle struc-

---

(*) Par exemple , le programme d'analyse de P.B.Braun, J.Hornstra, et J.I.Leenhouts (2), a permis à ces chercheurs de publier dans le courant de l'année 1969, les structures de 9 stéroïdes naturels, dont une contenait 2 molécules indépendantes dans la maille asymétrique.
A cette époque les méthodes directes n'avaient permis la réalisation que de 2 structures stéroïdiques :
- la digitoxigenine de J.Karle (28)
- l'oestriol de Cooper, Norton et Hauptman (19)

ture déterminée modifiait notre conception initiale,que nous exposerons dans ce travail. En graduant les difficultés, nous accèderons finalement à des structures cristallines d'une relative complexité, et nous décrirons la stratégie qui nous semble la plus adaptée à la recherche de structure de composés biologiquement actifs.

Notre travail est illustré essentiellement par des résolutions de structure de dérivés stéroïdiques à propos desquels nous avons entrepris une étude pharmacologique sur la relation structure-activité. Au moment où nous avons commencé ce travail, le"stéroïde"semblait représenter pour le cristallographe un degré de difficulté relativement élevé,et une seule structure de stéroïde naturel (c'est à dire sans atome lourd) avait été déterminée par rayons X :l'androstérone de D.F.High (1). Cependant,le développement complet de ce travail a nécessité la détermination de quelques structures de types différents, et finalement nous pensons que le mode de résolution proposé peut s'appliquer à l'étude de n'importe quel dérivé actif.

PREMIERE PARTIE

PREMIERE APPROCHE DES METHODES DIRECTES

INTRODUCTION

Les méthodes directes, dans le cas des structures non centrées, peuvent être divisées en trois grands groupes :

- l'addition symbolique des phases.
- la technique de la multisolution.
- le calcul direct des phases invariantes.

Nous avons voulu, dans un premier temps, utiliser chacun de ces modes de résolution pour étudier des structures de composés stéroïdiques, correspondant à des difficultés de même importance. Ce travail expérimental nous a permis de mettre au point quelques règles stratégiques que nous présentons dans cette première partie. Cependant il nous semble utile de faire, tout d'abord, un rappel sur les fondements théoriques de la recherche de structure par méthodes directes.

CHAPITRE I

GENERALITES SUR LES METHODES DIRECTES

I-1 POSITION DU PROBLEME

La fonction de répartition électronique relative à un atome de nombre atomique Z_j, situé à l'extrémité du vecteur r_j et supposé agité isotropiquement s'écrit :

$$\rho_j(r) = k\, Z_j\, e^{-k'(r-r_j)^2}$$

où k et k' sont des constantes

Par la suite, la densité électronique, dans la maille cristalline, peut se mettre sous la forme :

$$\rho(r) = k \sum_{j=1}^{N} Z_j\, e^{-k'(r-r_j)^2}$$

où N est le nombre d'atomes dans la maille cristalline.

En première approximation, et en admettant les atomes répartis au hasard, la position de chacun d'eux est définie par la connaissance des 3 coordonnées du vecteur r_j (x_j, y_j, z_j). Dans ces conditions, la connaissance de la structure $\rho(r)$ nécessite la détermination de 3N paramètres dans le groupe de symétrie P1, ou de $\frac{3N}{s}$ dans un groupe quelconque où le degré de symétrie est s.

Si on ne s'intéresse qu'aux atomes semi-lourds (carbone, azote, oxygène), pour une structure de 20 atomes, la densité électronique dans la maille dépend donc au maximum de 60 paramètres.

Or la densité électronique peut être calculée à partir des facteurs de structure par transformation de Fourier :

$$\rho(r) = \sum_H F_H \, e^{-i2\pi Hr}$$

où H désigne le vecteur de l'espace réciproque de composantes h, k, l.

Pour une structure de l'importance de celle que nous avons considérée, il est possible de déterminer environ 2000 facteurs de structure. Même si on tient compte du fait que l'expérience ne donne que les modules de ces facteurs, on dispose pour les 60 paramètres inconnus du problème, de 2000 relations du type :

$$F_H = \left| T.F_{(H)} \sum_{j=1}^{N} Z_j \, e^{-k'(r-r_j)^2} \right|$$

On a donc un système surabondant: il doit exister, entre les différents facteurs de structure F_H, des relations permettant de le résoudre mathématiquement.

I-2 MODE GENERAL DE RESOLUTION

L'équation de Sayre, établie en 1952 (31), est devenue l'équation de base des méthodes directes ; elle lie chaque facteur de structure à l'ensemble des autres facteurs de structure de l'espace :

$$F_H = k \sum_K F_K \, F_{H-K}$$

En réalité, le nombre de facteurs de structure est limité par la dimension de la sphère de résolution, et de plus, on ne considère généralement que les facteurs de structure les plus importants. Si les 3 termes F_H, F_K et F_{H-K} sont importants, on peut admettre avec une bonne probabilité que la relation $F_H \simeq k' \, F_K \, F_{H-K}$ reste vraie.

En pratique, au lieu du facteur de structure F_H, on préfère utiliser le facteur de structure normalisé E_H, qui est corrigé de l'agitation thermique.

$$E_H = \frac{1}{\sqrt{\sigma_2}} \sum_{j=1}^{N} Z_j \, e^{i2\pi Hr_j}$$

$$\text{où} \quad \sigma_2 = \sum_{j=1}^{N} Z_j^2$$

Le module du facteur de structure normalisé peut aussi être calculé à partir du module du facteur de structure observé F_H :

$$|E_H|^2 = \frac{|F_H|^2}{\varepsilon \sum_{j=1}^{N} f_j^2(H)}$$

où $f_j(H)$ est le facteur de diffusion atomique

ε un poids tenant compte des extinctions systématiques.

La relation $E_H \simeq E_K E_{H-K}$ correspondante est appelée relation $\sum_2$ de Karle et Hauptman.

Si les différents atomes de la maille sont repérés par rapport à un atome particulier de l'espace (x_o, y_o, z_o), on peut facilement écrire le facteur de structure normalisé sous la forme :

$$E = e^{i2\pi(hx_o + ky_o + lz_o)} \times \varepsilon_H = e^{i\Theta_{OH}} \varepsilon_H$$

où ε_H ne dépend que de la répartition des atomes les uns par rapport aux autres. Par la suite le module $|E_H| = |\varepsilon_H|$ ne dépend également que de cette répartition. Il n'en est, par contre, pas de même de sa phase Φ_H, qui dépend des paramètres de translation x_o, y_o, z_o localisant la molécule considérée par rapport à l'origine de la maille cristalline.

Or, dans les différents groupes de symétrie plusieurs points de l'espace peuvent servir d'origine; dans le groupe P1, sa localisation est même complètement arbitraire.

On appelle "invariant de structure" une grandeur qui ne dépend pas du choix de l'origine. ε_H , $|E_H|$ sont des invariants de structure.

Hauptman a montré (25) que certaines phases ou combinaisons linéaires de phases pouvaient également constituer des invariants de structure. La combinaison la plus simple est celle qui lie 3 réflexions H_1, H_2 , H_3 telles que :

$$H_1 + H_2 + H_3 = 0$$

Dans ce cas, le fait que $\Theta_{H_3} = -(\Theta_{H_2} + \Theta_{H_1})$ entraine que le triple produit $E_{H_1} E_{H_2} E_{H_3} = \varepsilon_{H_1} \varepsilon_{H_2} \varepsilon_{H_3}$ ne dépend pas du choix de

l'origine. Il en est de même de sa phase :

$$\Phi_{H_1} + \Phi_{H_2} + \Phi_{H_3} = \alpha$$

Par la suite, nous ne désignerons plus sous le nom d'invariants que cette somme $\Phi_{H_1} + \Phi_{H_2} + \Phi_{H_3}$ et naturellement les "cosinus et sinus invariants" en seront les cosinus et sinus (*) .

L'utilisation de la relation $\sum_2$ de Karle et Hauptman revient à considérer que toutes les phases invariantes sont nulles, puisque :

$|E_H|\, e^{i\Phi_N} = |E_K|\, e^{i\Phi_K} \times |E_{H-K}|\, e^{i\Phi_{H-K}}$ entraine $\Phi_H = \Phi_K + \Phi_{H-K}$

Les différents invariants ne sont pas indépendants les uns des autres. Vaughan (33) a pu décrire des formes "quadripolaires" :

$$\begin{aligned}
\Phi_{H_1} + \Phi_{H_2} + \Phi_{\bar{H}_1+\bar{H}_2} &= \alpha \\
\Phi_{\bar{H}_1} + \Phi_{H_3} + \Phi_{H_1-H_3} &= \beta \\
\Phi_{\bar{H}_2} + \Phi_{\bar{H}_3} + \Phi_{H_2+H_3} &= \gamma \\
\Phi_{\bar{H}_2+\bar{H}_3} + \Phi_{H_3-H_1} + \Phi_{H_1+H_2} &= \delta
\end{aligned}$$

où l'on voit aisément que la somme des 4 invariants constitutifs du quadripole ($\alpha+\beta+\gamma+\delta$) est identiquement nulle.

L'origine de la maille est fixée par le choix arbitraire de la phase d'un certain nombre (au maximum 3) de réflexions linéairement indépendantes, dans l'espace réciproque (25)(nous les désignerons par la suite sous le nom de réflexions "origine"). Toutes les phases de réflexions de l'espace sont déterminées de façon unique :

si dans le groupe de symétrie P222 on a fixé la phase d'une réflexion uug, les phases de toutes les autres réflexions uug seront déterminées par des invariants du type :

$$\Phi_{uug} + \Phi_{u'u'g'} + \Phi_{-(u+u',\ u+u',\ g+g')} = \alpha$$

où la troisième phase appartenant à la classe ggg est un invariant de structure et peut théoriquement être calculée à priori.

(*) Notons que la phase de certaines réflexions constitue aussi un invariant de structure. Par exemple dans les groupes du type P222 toute phase Φ_{ggg} est un invariant de structure.

En réalité, dans les groupes de symétrie ne contenant pas d'éléments d'inversion (centre ou miroir) le choix des phases des réflexions "origine" ne suffit pas pour déterminer toutes les phases des autres réflexions de l'espace de manière unique.

Pour les groupes P1, P2, P222,... il n'existe dans le cristal qu'un seul isomère optique (d) ou (l) mais il est à priori impossible de savoir lequel. En plus de l'origine, il est donc nécessaire de choisir arbitrairement l'un des 2 énanthiomorphes possibles.

Dans le groupe P222, le facteur de structure comprend un terme réel et un terme imaginaire :

$$A_c = \sum_j \cos 2\pi hx_j \cos 2\pi ky_j \cos 2\pi lz_j$$

$$\text{et} \quad B_c = \sum_j \sin 2\pi hx_j \sin 2\pi ky_j \sin 2\pi lz_j$$

Sa phase est égale à Arc tg B/A . Nous avons vu que la valeur absolue d'une phase Φ_{ggg} était un invariant de structure et ne dépendait, dès lors, que des modules des facteurs de structure. Par contre son signe, qui est aussi celui de $B_{ggg} \neq 0$ dépend de l'énanthiomorphe choisi : en effet changer d'énanthiomorphe revient à changer de signe toutes coordonnées atomiques donc à changer le signe de B_{ggg} . On peut faire choix d'un énanthiomorphe, en fixant arbitrairement le signe d'une phase Φ_{ggg} , pourvu qu'elle soit différente de 0 ou π (car $\pi = -\pi$ sur le cercle trigonométrique).

A partir des phases fixées ainsi arbitrairement, il est possible à l'aide de la relation $\sum_2$ ou des invariants de structure de déterminer, de proche en proche, les phases de toutes les réflexions de l'espace.

A la valeur $\Phi_H = \Phi_K + \Phi_{H-K}$ est associée la probabilité définie par Cochran, pour une structure centrosymétrique (16).

$$P_+ = \frac{1}{2} + \frac{1}{2} \text{ th } \frac{\sigma_3}{\sigma_2^{3/2}} |E_H E_K E_{H-K}|$$

Nous définirons, au chapitre suivant, la notion de probabilité associée à une phase quelconque dans le cas d'une structure non centrée. Par la suite nous désignerons le terme :

$$2 \frac{\sigma_3}{\sigma_2^{3/2}} |E_H E_K E_{H-K}| \quad \text{par} \quad A_{H,K}$$

I-3 LES FORMULES DE LA TANGENTE (29)

Parallèlement aux méthodes "algébriques" basées sur l'équation de Sayre ou la relation $\sum_2$ de Karle et Hauptman, se sont développées des méthodes de déterminations statistiques.

La valeur de l'invariant $\Phi_H + \Phi_K + \Phi_{H-K}$ n'est généralement pas nulle. Sa valeur dépend uniquement des modules des facteurs de structure ; elle est indépendante des phases de ces facteurs.

Une étude statistique portant sur l'ensemble des modules des facteurs normalisés de l'espace, permet de montrer notamment que (*) :

$$< \sin (\Phi_H + \Phi_K + \Phi_{H-K}) / A_{H,K} >_K = 0 \qquad (1)$$

$$\text{et que} \quad < \cos (\Phi_H + \Phi_K + \Phi_{H-K}) / A_{H,K} >_K = \frac{I_1(A_{H,K})}{I_0(A_{H,K})}$$

(*) *Si R_o et Φ_o d'une part, et R_1 et Φ_1 d'autre part désignent respectivement le module du facteur normalisé et la phase des réflexions K et H-K, pour une réflexion H fixée de l'espace réciproque, la densité de probabilité pour que $R_1 < |E_{H-K}| < R_1+dR_1$; $R_o<|E_K|< R_o+dR_o$; $\Phi_1< \Phi_{H-K} < \Phi_1 + d\Phi_1$ et $\Phi_o< \Phi_K < \Phi_o+d\Phi_o$ peut s'écrire ( ) :*

$$P(R_o,R_1,\Phi_o,\Phi_1) = C \exp \left[\frac{1}{1 - \frac{|E_H|^2 \sigma_3^2}{\sigma_2^3}} \left[-R_o^2 - R_1^2 + 2R_1R_o|E_H| \frac{\sigma_3}{\sigma_2^{3/2}} \cos(\Phi_H+\Phi_o-\Phi_1) \right] \right]$$

si on néglige les termes correctifs secondaires (C est une constante de normalisation).

Si on s'intéresse uniquement à la distribution de Φ_o et Φ_1, en supposant R_o et R_1 connus, cette expression prend la forme précédemment décrite par Cochran (16).

$$P(\Phi_o,\Phi_1) = C' \, e^{A_{H,K} \cos (\Phi_H + \Phi_o - \Phi_1)}$$

(on a négligé le terme $(\frac{\sigma_3}{\sigma_2^{3/2}})^2 |E_H|^2$ devant 1)

C' est alors égal à $\frac{1}{2\pi I_o(A_{H,K})}$ où I est la fonction de Bessel modifiée.

Il est à ce moment possible d'évaluer la moyenne d'une fonction $f(\Phi_o+\Phi_1)$ à l'aide de l'intégrale

$$\int P(\Phi_o+\Phi_1) \, f(\Phi_o+\Phi_1) \, d(\Phi_o+\Phi_1)$$

et notamment de calculer les valeurs moyennes des cosinus et sinus invariants. Pour cela, on utilise le développement en série de fonctions modifiées de Bessel I_n :

$$e^{A_{H,K} \cos(\Phi_H + \Phi_o - \Phi_1)} = I_o(A_{H,K}) + 2 \sum_{n=1}^{\infty} I_n (A_{H,K}) \cos n (\Phi_H + \Phi_o - \Phi_1)$$

Quand on calcule la valeur moyenne du cosinus, seul le terme n = 1 donnera une intégrale non nulle (car $<\cos^2\Theta> = \pi$); par contre dans le calcul de la moyenne du sinus tous les termes sont nuls.

Ces expressions sont des valeurs moyennes conditionnelles pour $A_{H,K}$ fixés (I_1 et I_o désignent les fonctions modifiées de Bessel).

Si on suppose H fixé, l'équation (1) peut être développée sous la forme :

$$\sin \Phi_H \cdot \langle \cos(\Phi_K + \Phi_{H-K}) / A_{H,K} \rangle_K + \cos \Phi_H \cdot \langle \sin(\Phi_K + \Phi_{H-K}) / A_{H,K} \rangle_K = 0$$

$$\text{d'où} \quad \operatorname{tg} \Phi_H = \frac{\langle \sin(\Phi_K + \Phi_{H-K}) / A_{H,K} \rangle_K}{\langle \cos(\Phi_K + \Phi_{H-K}) / A_{H,K} \rangle_K}$$

Si l'on veut s'affranchir de la dépendance de $A_{H,K}$, il est nécessaire de faire rentrer ce terme connu comme facteur multiplicatif dans les moyennes, (en réalité il serait préférable d'utiliser le poids $W_{\sin}(A_{H,K})$ associé $\langle \sin(\Phi_K + \Phi_{H-K}) \rangle$ pour A_{H-K}) ; on peut écrire :

$$\operatorname{tg} \Phi_H = \frac{\langle |E_K E_{H-K}| \sin(\Phi_K + \Phi_{H-K}) \rangle_K}{\langle |E_K E_{H-K}| \cos(\Phi_K + \Phi_{H-K}) \rangle_K} \qquad (3)$$

Cette relation que l'on désignera, par la suite, sous le nom de "formule de la tangente", n'utilise que partiellement les résultats de l'analyse statistique, car elle ignore la valeur moyenne du cosinus. Il en résulte un certain nombre d'imperfections et notamment une tendance à augmenter implicitement le nombre de phases invariantes égalées à zéro. Dans les chapitres suivants, nous reviendrons plus en détail sur cet inconvénient.

Il est possible d'utiliser simultanément les équations (1) et (2) en leur attribuant des poids $W_{\sin}$ et $W_{\cos}$ égaux à l'inverse de l'écart-type des lois de distribution des fonctions sinus et cosinus. En désignant par C_K le terme $\cos(\Phi_K + \Phi_{H-K})$ et par S_K le sinus correspondant, la combinaison linéaire de ces 2 équations :

$$W_{\sin} \, x(1) + W_{\cos} x(2)$$

ou plutôt le moindre carré construit à partir de (1) et (2), permet de définir la nouvelle phase Φ_H en minimisant l'expression :

$$\langle W_{\sin}^2 (C_K \sin\Phi_H + S_K \cos\Phi_H)^2 + W_{\cos}^2 \left(C_K \cos\Phi_H - S_K \sin\Phi_H - \frac{I_1(A)}{I_o(A)}\right)^2 \rangle_K \qquad (4)$$

A partir d'une évaluation approximative de Φ_H à l'aide de (3), d'utilisation plus commode, on pourra obtenir par (4) une solution affinée Φ_H plus proche de la valeur réelle.

Par analogie, on appelle (4) la "formule modifiée de la tangente".

C H A P I T R E II

METHODE D'ADDITION SYMBOLIQUE DES PHASES

Ce mode de recherche a été développée par J.Karle à partir de 1966 en utilisant pour les structures non centrées, une technique analogue à celle qu'il avait proposée autrefois pour les structures centrosymétriques et qui a depuis largement prouvé son efficacité. Cependant son utilisation dans le cas de structure non centrée est beaucoup plus délicate. Il est, dès lors,très difficile d'automatiser complètement cette méthode car de nombreuses interventions du "manipulateur" (particulières à chaque détermination) sont nécessaires pour assurer la convergence du processus vers une solution correcte.

Nous avons essayé de développer certaines variantes qui, sans porter atteinte à l'efficacité de la détermination "intuitive", permettent une automatisation relative du processus.

II-1 STRATEGIE INITIALE (J.Karle) (30)

A partir des 3 phases "origine" et au besoin de celle qui sert à fixer l'énanthiomorphe, on génère de proche en proche les phases des différentes réflexions de l'espace qui ont un facteur de structure normalisé

élevé (E > 1,5 ou 1,8),en utilisant la relation $\sum_2$ de Karle et Hauptman sous la forme :

$$\Phi_H = \Phi_K + \Phi_{H-K}$$

et en la supposant vraie avec une probabilité définie par :

$$P = \frac{1}{2} + \frac{1}{2} \text{ th } \frac{A_{H,K}}{2}$$

Comme nous le reverrons dans la IIème partie, le choix des réflexions "origines" n'est généralement pas suffisant pour générer les phases de tout l'espace. Il est nécessaire pour amorcer cette génération,d'attribuer à un certain nombre de réflexions de l'espace ayant un facteur E élevé, une phase symbolique "a" , "b"... que l'on utilise comme si elle était connue, bien que sa valeur numérique ne puisse être précisée qu'en fin de calcul, quand toutes les réflexions "permises (E > 1,5 ou 1,8) sont dotées d'une phase symbolique ou non.

Quand une même réflexion reçoit deux déterminations symboliques différentes, il est impossible de lui assigner une phase moyenne. Une phase n'est en effet déterminée qu'à 2 π près. Ne connaissant pas la valeur réelle de "a" et "b" on ne sait pas si la phase moyenne doit être calculée, pour obtenir la variance angulaire la plus faible, à partir des déterminations a et b, ou a et b+2π, ou a et b-2π... Si chacune des deux déterminations est forte la relation a=b (module 2π) sera prise en considération dans la suite du calcul. Si l'une des déterminations est faible, elle sera ignorée et la plus forte seule est prise en considération. (Nous discuterons plus loin les inconvénients de cette option).

Quand toutes les réflexions permises sont dotées d'une phase, on récapitule toutes les relations entre symboles et l'on essaie de résoudre le système d'équations obtenues. (Cette résolution est rendue difficile par le fait que chaque équation n'est définie qu'à 2π près ; si le nombre de symboles est supérieur à 2 ou 3 le problème devient rapidement inextricable).

Les 150 ou 200 phases dont on dispose alors, sont déterminées avec une certaine incertitude. Il est nécessaire de les affiner en utilisant la formule de la tangente et de générer d'autres phases pour les réflexions ayant un E > 1,4, puis à 1,2, et finalement E > 1.

A une détermination Φ_H donnée par la formule de la tangente on associe une grandeur α_H qui va permettre d'estimer la précision de ce résultat :

$$\alpha_H = \left| \left(\sum_K A_{H,K} \cos (\Phi_K + \Phi_{H-K})\right)^2 + \left(\sum_K A_{H,K} \sin (\Phi_K + \Phi_{H-K})\right)^2 \right|$$

La probabilité pour que la phase réelle de H soit Φ_{o_H} différent de Φ_H, a été établie par Cochran (16) :

$$P(\Phi_{o_H}) = \frac{1}{2\pi I_o(\alpha_H)} \; e^{\alpha_H \cos (\Phi_{o_H} - \Phi_H)}$$

où $I_o(\Phi_H)$ est la fonction de Bessel modifiée à argument imaginaire.

A partir de cette probabilité il est possible d'évaluer la variance :

$$Var\ (\Phi_H) = < (\Phi_{o_H} - \Phi_H)^2 >_K = <\Phi^2_{o_H}> - \Phi^2_H$$

où

$$< \Phi^2_{o_H} > = \frac{1}{2\pi I_o(\Phi_H)} \int_{\Phi_{H-\pi}}^{\Phi_H+\pi} \Phi^2_{o_H}\ e^{\alpha_H \cos(\Phi_{o_H} - \Phi_H)} d\Phi_H$$

En posant $x = \Phi_{o_H} - \Phi_H$ et en utilisant le développement en fonctions de Bessel :

$$e^{\alpha \cos x} = I_o(\alpha) + 2 \sum_{n=1}^{\infty} I_{2n}(\alpha) \cos 2nx + 2 \sum_{n=o}^{\infty} I_{2n+1}(\alpha) \cos(2n+1)x$$

on obtient finalement pour $Var\ (\Phi_H)$ l'expression :

$$Var\ (\Phi_H) = \frac{\pi^2}{3} + \frac{1}{I_o(\alpha)} \sum_{n=1}^{+\infty} \frac{I_{2n}(\alpha)}{n^2} + \frac{4}{I_o(\alpha)} \sum_{n=0}^{+\infty} \frac{I_{2n+1}(\alpha)}{(2n+1)^2}$$

Pour $\alpha_H = 10$ l'écart type correspondant est faible (19°) et la valeur trouvée est suffisamment précise pour assurer la convergence du calcul vers une solution correcte. Malheureusement on est souvent obligé d'utiliser une valeur $\alpha_{Hmin.}$ de 4.0 et même de 3. pour lequel l'écart type est égale à 38°).

Au cours de la génération de phases on évalue pour chaque réflexion, un facteur normalisé calculé statistiquement :

$$E_{H_{calc.}} = k \left(\frac{\sigma_2}{\sigma_3}\right)^{3/2} \left| <|E_K E_{H-K}| \cos(\Phi_K + \Phi_{H-K}) >^2_K + < |E_K E_{H-K}| \sin(\Phi_K + \Phi_{H-K}) >^2_K \right|$$

(k étant un facteur d'échelle)

On pourra, en fin de calcul, estimer l'exactitude de la génération de phases en considérant le facteur résiduel de Karle :

$$R_{Karle} = \frac{\sum_H |E_{Hobs.}| - |E_{Hcalc.}|}{\sum_H |E_{Hobs.}|}$$

Si les phases déterminées forment un ensemble cohérent, ce facteur est faible.

II-2 REALISATION DE PROGRAMMES D'ADDITION SYMBOLIQUE POUR IBM-1620

Pour réaliser ce processus d'Addition Symbolique, nous avons écrit un ensemble de 3 programmes permettant successivement de :

- calculer pour chaque réflexion les éléments nécessaires au stockage et le nombre d'interactions (il intervient au moment du choix de l'origine et des symboles).

- évaluer, pour les 200 plus fortes réflexions, une phase symbolique : celle-ci sera admise si la valeur α correspondante est supérieure à un seuil minimum. Si une réflexion admet plusieurs déterminations, celle qui est la plus forte lui est attribuée, les autres étant automatiquement considérées comme opposées et leurs contributions retranchées de la contribution la plus forte : ce sont des incompatibilités victuelles, qui disparaissent si la relation entre symbole correspondante est exacte.

- calculer la valeur numérique de la phase par la formule de la tangente, le facteur E_H calc. et le facteur résiduel de Karle.

Ce dernier programme, ainsi que le premier, recherche les différents couples de réflexions K et K-H correspondant à une réflexion H, en utilisant la notion de vecteur espace que nous avons déjà décrite par ailleurs (1968) .

Chaque réflexion H(d'indices h, k, l) est associé de façon biunivoque à une adresse de la mémoire centrale de l'ordinateur par la relation :

$$A_H = A_{(hkl)} = (2h_{max}+1)\,(2k_{max}+1)\,\ell + (2h_{max}+1)\,k + h + A_o$$

(pour le système triclinique)

Si on considère les adresses associées aux réflexions H et K, l'adresse de la réflexion K-H se calcule très simplement par :

$$A_{K-H} = (A_K - A_H) + A_o$$

L'utilisation de l'espace mémoire permet donc un accès direct (sans comparaison d'indices) aux éléments entrant dans une relation $\sum_2$.

II-3 APPLICATION AUX STRUCTURES "PSEUDOCENTREES" - RECHERCHE D'UNE STRATEGIE EXPERIMENTALE

(Nous appelons structures"pseudo centrées" , les structures appartenant aux groupes de symétrie non centrés de type P222 pour lesquelles les trois plans principaux sont centrés).

Après une tentative de résolution par addition symbolique, la structure de l'oestradiol hémihydraté a pu être déterminé par minimisation directe du facteur résiduel. Reprenant l'étude de cette structure par méthodes directes nous avons pu définir les erreurs qui avaient causé l'échec

initial et ainsi mettre au point les premières règles stratégiques qui nous ont permis d'entreprendre avec succès l'étude de la structure de l'oestrone II.

II-3-1 Analyse de la génération des phases pour l'oestradiol ($P2_12_12$)

D'après les règles établies par H.Hauptman (25), l'origine est fixée dans le groupe P222 par le choix des phases de trois réflexions indépendantes (c'est à dire de 3 réflexions n'appartenant pas à la classe ggg et dont les différentes combinaisons 2 à 2, ou 3 à 3 ne donnent pas ggg).

Généralement, pour les groupes pseudo centrés on choisit ces 3 réflexions dans les plans principaux, car les phases correspondantes sont situées sur les "poles" (c'est à dire ont pour seules valeurs possibles 0, π ou $\pm \pi/2$). Nous avons donc choisi une phase arbitraire pour les réflexions :

970	E = 3,27 π	uug
610	E = 2,47 0	gug
051	E = 2,32 $\pi/2$	guu

L'énanthiomorphe est ensuite défini par le choix du signe d'une phase par exemple la phase d'une réflexion ggg à condition qu'elle soit différente de 0 ou π . (On ne peut déterminer l'énanthiomorphe en choisissant la valeur $+\pi$, ou $-\pi$, ces deux valeurs correspondent en fait à la même phase). Comme on n'est jamais sûr à priori qu'une phase Φ_{ggg} diffère de 0 ou π, nous avons préféré utiliser la deuxième façon de fixer l'énanthiomorphe en considérant une combinaison de phases linéairement dépendantes que l'on construit égale à $\pm \pi/2$.

Si nous considérons la réflexion 041, elle forme avec les 2 réflexions origine 610 et 051 une combinaison linéairement dépendante. (Nous rencontrons ici un autre type d'invariants de structure constitué d'un ensemble de réflexions pour lesquelles la somme sur chacun des indices est paire). La phase de la réflexion 041 est 0 ou π. L'invariant considéré a pour phase :

$$\Phi_{610} + \Phi_{051} + \Phi_{041} = \pm \pi/2$$

On peut faire choix de l'un des énanthiomorphes possibles en fixant, pour cette combinaison de phases, la valeur $+\frac{\pi}{2}$, c'est à dire en choisissant :

041 E = 1,98 π

Compte tenu des réflexions choisies pour fixer l'origine et l'énanthiomorphe, nous avons été obligés d'introduire 2 symboles pour générer les phases de l'espace :

10	1	2	E = 2,57	a
2	10	4	E = 2,43	b

A chaque itération les réflexions pour lesquelles les phases étaient déterminées avec une probabilité satisfaisante (nous avions pris comme seuil $\alpha_H = 2{,}30$) étaient admises comme connues et jointes aux réflexions précédemment déterminées.

Une fois obtenues les phases de 76 réflexions ayant une valeur E > 1,50, les relations entre symboles permirent de trouver a = 0 et b = 0

Les valeurs réelles de ces deux symboles sont respectivement a = -35° et b = +9° ; les valeurs trouvées se situaient donc sur les pôles les plus proches. Cependant,l'affinement par la formule de la tangente entrepris à ce stade, ne converge,dans ce cas, vers aucune solution cohérente et près de 50 % des phases des réflexions ayant une valeur E > 1. différent de la valeur réelle de plus de 90°.

Remarquons que pour les 76 phases de départ, l'écart moyen entre la phase calculée et la phase réelle est déjà égal à 38° et pour 6 de ces réflexions cet écart est supérieur à 90°. L'utilisation des symboles revient à faire pour chaque détermination des approximations de plus en plus importantes,et plus le nombre d'itérations nécessaires pour fixer les valeurs des symboles est grand, plus les phases ainsi définies sont entachées d'erreur. Même, si à ce stade on avait attribué aux 2 symboles leur valeur réelle, la formule de la tangente n'aurait pas converger vers la bonne solution.

II-3-2 Première Stratégie expérimentale

A partir de cette étude nous avons pu établir :

- qu'il est préférable d'affecter des symboles à des réflexions pour lesquelles la phase peut être déterminée de façon précise, c'est à dire des réflexions appartenant aux plans principaux. Les phases de ces réflexions ne peuvent prendre que les valeurs $0, \pi$ ou $\pm \pi/2$. Pour plus de sureté on peut choisir des invariants de structure ggo, gog et ogg pour lesquels l'utilisation de la relation $\sum_1$ (voir paragraphe II-3-3) peut donner une indication sur la valeur de la phase.(Nous verrons au chapitre suivant que la multisolution fournit un moyen différent de corriger des relations légèrement fausses entre les symboles).

- qu'il est illusoire de déterminer le plus de phases possibles en abaissant le seuil $E_{min.}$,car pour des réflexions ayant une valeur E < 1,5 la marge d'erreur sur la phase calculée devient considérable, et,alors que le temps de calcul augmente de façon importante, la précision obtenue diminue au contraire notablement.

- enfin et surtout qu'il faut utiliser l'affinement par la formule de la tangente dès qu'apparait la première phase déterminée sur 2 pôles en quadrature c'est à dire à la fois égale à 0 ou π et à $\pm \pi/2$. Dans ce cas, la formule de la tangente donnera pour cette phase, une valeur proche de la valeur réelle. Au contraire si on utilise pour cette phase, même de façon provisoire, la valeur du pôle correspondant à la détermination la plus probable, elle sera encore renforcée par les interactions ultérieures et la phase obtenue en fin de calcul est entachée d'une erreur de l'ordre de 45° : aucune convergence vers une solution exacte n'est plus possible.

II-3-3 Application à l'oestrone II ($C_{18}\ O_2\ H_{20}$) (8)

Dans un solvant polaire l'oestrone cristallise dans le groupe spatial $P2_12_12_1$ avec une molécule dans l'unité asymétrique.

L'origine a été déterminée en fixant arbitrairement la phase de 3 réflexions linéairement indépendantes :

5 12 0	E = 3,51	$+\pi/2$
6 11 0	= 3,34	0
0 10 5	= 2,91	0

La réflexion 095 donne, avec les 2 réflexions origine 6 11 0 et 0 10 5 un invariant de structure dont la valeur est $+\frac{\pi}{2}$. Nous pouvons donc choisir l'un des énanthiomorphes en fixant la phase :

0 9 5	E = 2,01	$+\pi/2$

Pour amorcer la génération des phases, il est également nécessaire d'utiliser 2 symboles :

8 8 0	E = 2,73	a
0 10 4	= 2,13	b

La valeur de ces 2 symboles peut être prévue par l'utilisation de la relation $\sum_1$ de Karle et Hauptman (29b) pour le groupe $P2_12_12_1$

$$s_{2h,2k,0} = s \left| \sum (-1)^{h+\ell} \left(|E_{hk\ell}|^2 - 1\right) \right|$$

associée à la probabilité :

$$P_{\sum_1} = \frac{1}{2} + \frac{1}{2} \text{ th } C \left| \sum_{\ell} (-1)^{h+\ell} \left(|E_{hk\ell}|^2 - 1\right) |E_{2h,2k,0}| \right|$$

Pour la réflexion 880, $\sum_1 = +12,65$ on a donc a = 0 avec une probabilité $P_{\sum_1}$ pratiquement égale 1. Il en est de même pour la réflexion 0 10 4 , b = 0 avec $\sum_1 = +4,97$.

La génération symbolique des phases fut alors entreprise avec un seuil α_{Hmin} égal à 3,00 . Cependant, au bout de quelques itérations, il apparut pour certains invariants de structure ggo des phases opposées à celles données par la relation $\sum_1$:

2 2 0	π	avec	α = 34,34	alors que $\sum_1$	= + 3,83
4 12 0	π	"	α = 2,48	" "	= +13,14
6 10 0	π	"	α = 14,98	" "	= + 6,96

(Notons qu'une indication $\sum_1$, qui représente un calcul à priori, a généralement plus de valeur qu'une indication $\sum_2$ même avec une très forte probabilité car elle dépend des erreurs de déterminations commises précédemment ; les deux probabilités $\sum_1$ et $\sum_2$ n'ont pas la même signification).

En introduisant dans une nouvelle génération de phases ces 3 réflexions avec la valeur déterminée par $\sum_1$, nous avons pu mettre en évidence qu'au cours de l'étude précédente, il y avait eu 3 phases dont la détermination erronée se répercutait sur près de 15 % des réflexions :

1 11 5	E = 1,75	(+ π/2) déterminée	- π/2 au 2ème cycle avec	α = 3,00
1 11 4	E = 2,06	(+ π/2) "	- π/2+b au 3ème cycle avec	α = 3,72
4 3 6	E = 1,88	(π) "	π+b au 6ème cycle avec	α = 6,12

Ces 3 réflexions furent alors retirées et une nouvelle détermination entreprise. Il fut alors possible de fixer les 2 symboles par la relation $\sum_2$ conformément à la détermination obtenue par $\sum_1$ (a = 0 et b = 0)

Dès le 3ème cycle, il apparait une phase déterminée sur 2 pôles en quadratures :

6 8 2	E = 2,31	π	avec α = 2,80
		$-\frac{\pi}{2}$	α = 2,34

Nous avons donc commencé à ce stade l'affinement par la formule de la tangente avec, outre les 6 réflexions origine et symboles, 8 nouvelles réflexions :

8 2 4	E = 2,27	0°	1 9 5	E = 2,08	270°
6 1 5	2,76	180°	3 4 1	2,28	270°
5 2 5	2,07	270°	2 2 4	2,24	0°
3 14 4	1,74	90°	6 8 2	2,31	220°

(la valeur 220° a été calculée en prenant la moyenne des deux déterminations en quadrature

$$\frac{1}{\sum_K A_{H,K}} \left(\sum_K A_{H,K} \left(\Phi_K + \Phi_{H-K} \right) \right)$$

L'affinement fut poursuivi sur ces 14 réflexions jusqu'à ce que le facteur d'échelle $k = \frac{\sum E_{calc.}}{\sum E_{obs.}}$ demeure constant. De nouvelles réflexions sont alors déterminées. Pour que les différentes phases ne varient pas trop brusquement, nous n'avons introduit, à chaque cycle d'affinement, que 10 % de phases nouvelles. A chaque fois, l'affinement est poursuivi jusqu'à obtention d'un k stable.

Au bout de 4 heures d'ordinateur IBM-1620 (ce qui correspond à environ 8 mn. sur IBM-360-44) nous disposiions des phases de 142 réflexions ayant une valeur E > 1,5, c'est à dire environ 7 réflexions par atomes semi-lourds. La transformée de Fourier effectuée avec ces réflexions a permis de placer sans ambiguïté,malgré quelques pics parasites, 18 des 20 atomes semi-lourds et l'affinement de la structure a pu être réalisé sans difficultés (R_{final} = 0,039).

Le temps de calcul nécessaire pour cette détermination est relativement faible. Il reste cependant un point hasardeux :
on peut en effet considérer comme très heureux d'avoir pu "deviner" l'apparition de déterminations erronées dans les premiers cycles de calcul en utilisant les relations $\sum_1$. Si ces mauvaises déterminations n'avaient pas influé sur des invariants de structure nous aurions été incapables de les corriger.

Si l'on veut résoudre des structures plus complexes par addition symbolique il faut pouvoir éviter l'apparition de phases erronées ou du moins les déceler. Les procédés habituellement utilisés, changements d'origine, utilisation de phases connues..., ne peuvent être appliqués de façon systématique.

La technique que nous proposons maintenant, ne fait appel à aucune considération particulière; elle pourra être utilisée pour n'importe quelle détermination.

II-4 LE CONTROLE SYMBOLIQUE DES PHASES (4)

Dans l'étude d'une structure par la méthode d'addition symbolique, la partie la plus délicate est la détermination des 50 premières phases :

- c'est à ce moment que le nombre de phases erronées est le plus grand (parce que l'on ne dispose que de peu d'informations pour calculer une nouvelle phase).

- c'est à ce niveau qu'une détermination erronée a les conséquences les plus graves; l'erreur va s'étendre de proche en proche à un ensemble important de réflexions.

De ce fait, bien souvent, l'échec de l'addition symbolique est due à l'apparition d'invariants de structure supérieur à 60° dans les premiers cycles de calcul (donc d'une phase déterminée avec une erreur supérieure à 60°). Or si la probabilité, pour qu'un invariant prenne une telle valeur, est faible quand $A_{H,K}$ est supérieur à 4,5 (Prob < 0,05) (*) , elle

---

(*) H.Hauptman a montré (21) que la distribution conditionnelle du cosinus invariant, pour $A_{H,K}$ donné, peut être définie par la densité de probabilité :

$$(x) = \frac{\exp A_x}{\pi I_o(A)\sqrt{1-x^2}} \qquad si\ |x| < 1$$

où $I_o(A)$ est la fonction de Bessel à argument imaginaire

$$(x) = 0 \qquad si\ |x| > 1$$

Nous considérons ici la probabilité que ce cosinus soit compris entre -1 et +1/2 qui est égale à $1 - \int_{1/2}^{1} \rho(x)dx$, en fonction de $A_{H,K}$.

n'est plus négligeable si la valeur de $A_{H,K}$ diminue : pour A ≃ 2,0 elle devient égale à 0,22.

Si l'on admet, par exemple, que la méthode converge vers une solution correcte si les 10 premiers invariants utilisés sont inférieurs à 60°, la probabilité d'un tel évènement est égale à $(0,95)^{10} = 0,59$ dans le cas où $A_{H,K}$ est supérieur à 4,5. Elle n'est plus que $(0,78)^{10} = 0,08$ si $A_{H,K} \simeq 2,0$.

Or nous avons vu, pour l'oestrone II que l'apparition des phases erronées se produit bien plus tard au cours de la détermination. Elle empêche malgré tout la convergence de la méthode. Il est pratiquement impossible, par les techniques décrites précédemment, de ne pas faire apparaitre de déterminations erronées pour les 50 premières phases.

$A_{H,K}$ est une fonction de $\sigma^3/\sigma_2^{3/2}$ ou de $\frac{1}{\sqrt{N}}$ pour une structure à atomes identiques : il dépend de la complexité de la structure. Si on n'utilise au départ que des facteurs normalisés supérieurs à 2,5 (ce qui est déjà bien difficile à réaliser), une valeur $A_{H,K}$ supérieure à 4,5 correspond à un nombre N d'atomes dans la maille cristalline au plus égal à $\left(\frac{2E^3}{A}\right)^2 \simeq 48$; soit, dans le cas du groupe $P2_12_12_1$, (Z = 4) à une molécule de 12 atomes.

Pour atténuer les inconvénients qui résultent de l'apparition d'invariants de structure supérieurs à 60° et donc pour rendre efficace la méthode d'addition symbolique dans le cas de structures plus complexes, nous proposons la technique d'étude suivante :

→ Dans un premier stade nous essaierons d'augmenter pour chaque réflexions la précision de la détermination : au lieu d'accepter à chaque itération toutes les réflexions ayant une valeur α_H au moins égal à 3,0, c'est à dire pour lesquelles la probabilité d'avoir une erreur supérieure à 60° est égale à 0,11, nous n'accepterons que la réflexion dont la phase est la mieux déterminée. De cette façon, on augmente la valeur moyenne du seuil d'acceptation et on diminue donc les possibilités de déterminations de phases erronées.

→ A ce moment, nous pourrons admettre que les phases acceptées avec une valeur α_H supérieure ou égale à 8,0 sont déterminées de façon correcte (la probabilité pour que l'erreur commise soit supérieur à 60° est en effet inférieure à 0,6 %).

Par contre, nous admettrons que les phases déterminées avec une valeur α_H inférieure à 8,0 peuvent être erronées : l'assignation symbolique supposée fausse, est alors remplacée par un autre symbole qui servira à contrôler la première détermination. Si, à la fin de la génération "symbolique" de phases, cette assignation est trouvée absolument égale à celle du premier calcul, la phase douteuse sera confirmée. Si, au contraire, nous trouvons pour ce symbole de "contrôle" une indication différente ou des indications équiprobables nous pourrons estimer que la première assignation était erronée.

(Cette utilisation d'un symbole de contrôle peut paraître bien artificielle et, effectivement, nous l'avons considérée au début comme une fantaisie qui devait permettre de résoudre des structures ; par la suite, nous avons compris qu'il résultait de façon logique de l'existence d'ensemble de corrélations dans l'espace réciproque, nous les traiterons en détail dans la deuxième Partie. Le symbole de contrôle permet de tester si une réflexion appartient à la zone d'intersection de deux iles. Nous pensons donc qu'à la fin de cet ouvrage notre lecteur saisira mieux l'intérêt de cet artifice).

* *
*

Application à l'oestrone II ($P2_12_12_1$)

Si nous utilisons à chaque fois la phase la mieux déterminée pour poursuivre la génération, il n'y a plus que 2 réflexions acceptées avec une valeur α_H inférieure à 4,0 , sur les 42 réflexions premières déterminées.

1 9 5	E = 2,08	α = 3,14	4ème réflexion déterminée
3 14 4	E = 1,74	α = 3,02	5ème réflexion déterminée

Il y en avait 21 (soit 50 %) lors de la détermination initiale.

De plus, avec cette première modification, les réflexions 1 11 5 et 1 11 4 qui étaient erronées, disparaissent de l'ensemble des réflexions acceptées comme connues.

Par contre l'assignation erronée, obtenue pour la réflexion 4 3 6, persiste : elle est admise en 17ème position avec une valeur α_H égale à 7,36 , et si au début elle ne semble pas entrainer d'erreurs, à partir de la 32ème réflexion admise toutes les assignations deviennent erronées.

Nous avons utilisé ensuite un symbole de contrôle pour les différentes réflexions acceptées avec une valeur α_H inférieure à 8,0. Si la valeur finalement calculée pour ce symbole confirme la détermination initiale pour toutes les autres réflexions, le symbole affecté à la réflexion 4 3 6 apparait en fin de calcul avec deux déterminations contradictoires de même importance : π+a et a+b. Nous pouvons alors supposer que la détermination initiale π+a était erronée, et que la valeur exacte de la phase symbolique de la réflexion 4 3 6 est a+b.

Notons, en passant, que l'assignation d'un symbole de contrôle à la réflexion 4 3 6 fait disparaître, d'emblée, la venue d'assignation ultérieures fausses : les 42 premières réflexions acceptées ont toutes la bonne assignation symbolique quelle que soit la valeur du symbole de contrôle de la réflexion 4 3 6.

Dans la deuxième partie, nous verrons une application de cette notion à la détermination de la structure de la testostérone.

II-5 CONCLUSION

La relation $\sum_2$ de Karle et Hauptman (1953) présentait une telle simplicité que les cristallographes ont longtemps estimé pouvoir ignorer la notion de "symboles" que Gillis avait introduit à propos des équations de Harker et Kasper (1949). Mais la détermination de structure à l'aide de cette relation n'a pu être développée qu'à partir du moment où J.Karle reprit l'idée du symbole (1963) : l'utilisation de la relation $\sum_2$ est alors devenue la méthode d'addition symbolique.

Au cours de ce chapitre, nous avons vu que cette méthode s'avérait très efficace pour la détermination de structure de complexité moyenne; nous avons également essayé d'apporter quelques améliorations à la technique décrite par J.Karle.

Cependant les échecs de l'addition symbolique ont incité les cristallographes à rechercher de nouvelles méthodes de détermination, plus "puissantes" (multisolution, invariants de structure). Dans les chapitres suivants, nous étudierons expérimentalement ces différentes techniques de résolution.

Mais nous restons persuadé que le "symbole" demeure un remarquable outil de détermination et nous essaierons dans la 2ème Partie de justifier son emploi, à partir de l'interprétation physique des méthodes directes.

CHAPITRE III

METHODE DE LA MULTISOLUTION

Le développement des moyens de calcul a permis l'élaboration de techniques plus compliquées dans lesquelles on essaie de diminuer au maximum les possibilités d'erreurs. Dans la méthode de la multisolution on donne aux phases supposées mal déterminées ou aux phases indéterminées (symboles), toutes les valeurs possibles comprises entre 0 et 360° (en réalité on fait choix d'un certain nombre de déterminations situées à intervalles réguliers).

Il en résulte dans tous les cas, un grand nombre d'ensemble de départ à affiner par la formule de la tangente (en général plus de 64 solutions) nécessitant, soit un ordinateur de grande puissance, IBM-360-75, CDC6600,.. soit des programmes très élaborés. Nous considérerons plus spécialement le programme MULTAN de P.Main et G.Germain qui est actuellement un des plus efficaces et certainement le plus utilisé par les différents cristallographes.

Cependant, restant fidèle à l'addition symbolique, nous avons essayé de développer une technique où ce procédé serait associé à la multisolution. De cette façon on réduit considérablement le nombre de solutions à étudier (généralement moins de 8) sans pour autant, diminuer l'efficacité de la multisolution.

Pour trier les différentes solutions on fait appel à un certain nombre de figures de mérite ; nous consacrerons un paragraphe à une étude critique des différentes fonctions utilisées dans ce but. Enfin nous terminerons par un exemple d'application de l'utilisation simultanée de l'addition symbolique et de la multisolution.

III-1 STRATEGIE INITIALE (M.M.WOOLFSON, G.GERMAIN, P.MAIN) (22-23)

Si on utilise la relation de J.Karle qui exprime la variance d'une phase calculée en fonction de α_H (voir II-1), on peut évaluer l'écart-type $\sigma_{(\alpha)}$ associé à cette détermination. Il est alors possible d'estimer l'erreur avec laquelle cette phase a été calculée ; en admettant que cette erreur est distribuée suivant une gaussienne, on peut écrire au seuil de 95 %:

$$\Phi_H = \Phi_{calcH} \pm 1{,}96\ \sigma_{(\alpha)}$$

Comme nous l'avons déjà signalé, pour que la formule de la tangente fournisse une solution correcte, il ne faut pas que l'erreur commise sur chaque nouvelle phase déterminée soit supérieure à 60°. Ceci est réalisé si la valeur α_H associée à cette phase est suffisamment élevée. Par contre l'erreur croit rapidement quand la valeur α décroît : elle est supérieure à 90° si $\alpha_H = 2{,}3$.

La méthode de la multisolution va essayer de diminuer cette erreur ; au lieu d'attribuer à la réflexion H la phase Φ_{calcH} déterminée par la formule de la tangente et qui est, dans ce cas, entachée d'une erreur de 90°, on préfère entreprendre deux calculs distincts en prenant comme valeur de Φ_H respectivement $\Phi_{calcH} + 45°$ et $\Phi_{calcH} - 45°$. L'erreur commise sur chacune de ces réflexions n'est plus égale qu'à 45°. On recommence le processus chaque fois que l'on est amené à utiliser une réflexion déterminée avec une valeur α_H faible.

S'il existe n réflexions de ce type, on voit aisément que le nombre de solutions à étudier est 2 n. Il serait donc peut être plus judicieux d'éviter l'apparition d'un trop grand nombre de phases mal déterminées.

Ce but est en partie atteint quand on utilise le processus de"convergence" (P.Main (23)) pour choisir l'origine. A partir des règles fixées par H.Hauptman pour choisir les réflexions "origine", il est généralement possible de sélectionner différents groupes de 3 (si on est en mode P) réflexions dans l'ensemble des facteurs E_H utilisés. On tient compte généralement de l'importance du facteur normalisé et du nombre d'interactions (relation $\sum_2$). Mais un tel choix n'assure pas automatiquement, par la suite, la génération de phases avec une valeur α_H élevée.

La valeur α_H déterminée par Karle et utilisée dans la méthode d'addition symbolique, ne peut être calculée qu'une fois la phase Φ_{calcH} déterminée à partir de la formule de la tangente. Il est cependant possible de donner à priori une estimation de la valeur α_H en l'absence de toute information sur les phases Φ_H :

$$\alpha^2_{Hest} = \sum_{H'} A^2_{H,H'} + 2 \sum_{H'} \sum_{\neq H''} A_{H,H'} \times A_{H,H''} \frac{I_1(A_{H,H'})}{I_0(A_{H,H'})} \times \frac{I_1(A_{H,H''})}{I_0(A_{H,H''})}$$

où I_1 et I_0 désignent les fonctions modifiées de Bessel d'ordre 0 et 1 (*)

Ainsi, sans faire le calcul de phases, on peut, à partir de l'origine, calculer les différentes valeurs $\alpha_{Hest.}$. On veut choisir l'origine de façon que les valeurs $\alpha_{Hest.}$ générées de proche en proche soit les plus grandes possibles. Pour faire ce choix, il est plus facile de prendre le problème à l'envers. En éliminant chaque fois la réflexion correspondant à la valeur α_{Hest} la plus faible, les dernières réflexions restantes seront celles qu'il faut prendre comme origine.

(*) *Ce calcul utilise les mêmes éléments que le calcul de la variance, à partir de* α_H *calculé, reproduit au paragraphe II-1.*

Soit Φ_{oH} *la phase réelle (inconnue) d'une réflexion H, pour laquelle le couple* $K, H-K$ *donne une indication* $\Phi_H = \Phi_K + \Phi_{H-K}$. *La probabilité de Cochran pour le seul terme peut s'écrire :*

$$P(\Phi_{oH}) = \frac{1}{2\pi I_o(A_{H,K})} \; e^{A_{H,K} \cos(\Phi_{oH} - \Phi_H)}$$

Or le terme $\alpha^2_{H\ calc.}$ *s'écrit, en prenant pour origine des phases la valeur* Φ_{oH} *:*

$$\alpha^2_H = \left| \sum_K A_{H,K} \cos(\Phi_K + \Phi_{H-K} - \Phi_{oH}) \right|^2 + \left| \sum_K A_{H,K} \sin(\Phi_K + \Phi_{H-K} - \Phi_{oH}) \right|^2$$

Soit en développant les carrés :

$$\alpha^2_H = \sum_K A^2_{H,K} + \sum_K \sum_{\neq K'} A_{H,K} A_{H,K'} \cos(\Phi_{K'} + \Phi_{H-K'} - \Phi_{oH}) \cos(\Phi_K + \Phi_{H-K} - \Phi_{oH})$$

$$+ \sum_K \sum_{\neq K'} A_{H,K} A_{H,K'} \sin(\Phi_{K'} + \Phi_{H-K'} - \Phi_{oH}) \sin(\Phi_K + \Phi_{H-K} - \Phi_{OH})$$

Si on utilise alors les valeurs moyennes des sinus et cosinus invariants définies au paragraphe I.3. :

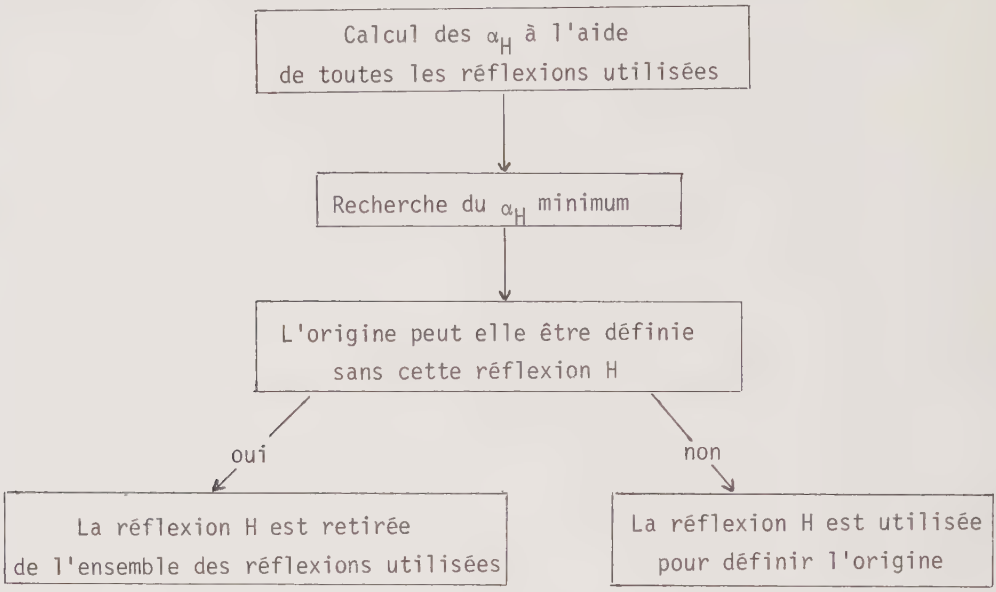
$$\langle \cos(\Phi_K + \Phi_{H-K} - \Phi_{oH}) \rangle_K = \frac{I_1(A_{H,K})}{I_o(A_{H,K})} \quad et \quad \langle \sin(\Phi_K + \Phi_{H-K} - \Phi_{oH}) \rangle_K = 0$$

pour évaluer les sommes sur K et K', on obtient l'expression utilisée ci-dessus.

Généralement, pour calculer le rapport $\frac{I_1(A)}{I_o(A)}$, *on utilise l'approximation :*

$$\frac{I_1(A)}{I_o(A)} \simeq 0{,}5658\ A - 0{,}1304\ A^2 + 0{,}0106\ A^3$$

(On peut résumer le processus de convergence par l'ordinogramme ci-dessous (P.Main (23)) :



On affecte systématiquement une multidétermination ($\pm \frac{\pi}{4}$ et $\pm \frac{3\pi}{4}$) aux réflexions qui, lors de leur élimination dans le processus de convergence, avait une valeur α_{est} inférieure à un seuil $\alpha_{min.}$ fixé à l'avance. Par analogie avec l'addition symbolique, nous continuerons à appeler "symbole" une réflexion multidéterminée.

III-2 UTILISATION SIMULTANEE DE LA MULTISOLUTION ET DE L'ADDITION SYMBOLIQUE

Par rapport à l'addition symbolique, la méthode de la multisolution présente de nombreuses améliorations, mais aussi quelques inconvénients. S'il est incontestable qu'en moyenne l'origine déterminée par le processus de convergence est meilleure que celle fixée, en quelque sorte au hasard, dans la méthode d'addition symbolique, il ne faut pas croire qu'elle sera toujours la meilleure : sa détermination dépend intimement de l'algorithme utilisé. Ainsi, une réflexion "origine" possible donnant peu d'interactions est rapidement éliminée par le processus de convergence ; elle peut cependant être fortement liée à une réflexion qui donne beaucoup d'interactions et constitue une "origine" valable. (En fait, la seule façon correcte de définir l'origine consiste à considérer toutes les "origines" permises par les règles d'Hauptman et à étudier leur aptitude à générer des réflexions fortement liées).

Chacun des symboles utilisés nécessite l'étude de 4 solutions différentes. Si le processus de convergence impose n symboles pour générer les phases de l'espace, il faudra étudier 4^n solutions différentes,

avec au besoin un nombre égal de sections de Fourier, avant d'obtenir une solution correcte. Or comme nous le verrons dans la 2ème partie, pour la plupart des groupes de symétrie, s'il n'y a qu'une seule molécule dans la maille asymétrique, il n'est théoriquement pas nécessaire d'introduire de symboles pour générer l'ensemble des phases de l'espace. Le nombre de symboles nécessaires augmente par contre dès que l'on trouve différentes molécules dans la maille asymétrique.

Généralement le processus de convergence définit plus de symboles qu'il n'est physiquement nécessaire. Comme le choix de l'origine, le nombre de symboles demandés dépend de l'algorithme de convergence utilisé. Il y a donc un certain nombre de symboles inutiles qui doivent pouvoir être déterminés par un ensemble de relations $\sum_2$ fortes.

Certains symboles peuvent donc être éliminés par l'utilisation de l'addition symbolique préalablement à la multisolution. Si le symbole correspond à une réflexion en position générale, on pourra admettre, dans l'esprit de la multisolution, que la détermination obtenue (par exemple $a = \frac{\pi}{2}$) est entachée d'une certaine erreur et étudier 2 solutions différentes ($a = + \frac{\pi}{2} \pm \pi/4$).

Si au départ le processus de convergence définit n symboles et si l'addition symbolique fait apparaitre m relations entre ces symboles, le nombre de solutions à étudier n'est égal qu'à 2^{2n-m} alors qu'il était égal à 2^{2n} dans le mode d'application initial de la multisolution.

Pour diminuer le nombre de symboles déterminés par le processus de convergence on peut également utiliser le processus décrit au paragraphe II.4 à partir de l'origine choisie. Chaque fois qu'aucune réflexion nouvelle ne peut plus être déterminée avec une valeur α satisfaisante, on introduit comme symbole la première réflexion rencontrée dans le sens inverse du processus de convergence et dont la phase n'a pas pu être déterminée.

Remarque : Si cette technique s'avère avantageuse dans l'étude des structures pseudo centrées, car elle diminue de façon notable le nombre de solutions à étudier, elle peut présenter certains dangers dans le cas de groupes monocliniques ou tricliniques non centrées; nous y reviendrons un peu plus loin.

III-3 LES FIGURES DE MERITE

Une fois généré, pour chacune des solutions de la multisolution, l'ensemble des phases des réflexions de l'espace ayant un facteur E supérieur à E_{min}, il faut déterminer laquelle ou lesquelles de ces solutions semblent les plus probables et sont susceptibles de donner des sections de densité utilisables. Dans ce but, on utilise un certain nombre de figures de mérite. Le programme initial de P.Main en utilise 4 différentes :

- le facteur résiduel de Karle

$$R_{Karle} = \frac{\sum_H |E_{calc}(H) - E_{obs}(H)|}{\sum_H |E_{obs}(H)|}$$

traduit l'autocohérence de l'ensemble de la génération de phases; il doit être minimum.

- la somme $S_\alpha = \sum_H \alpha_H$ doit être maximale; elle rend compte en quelque sorte de la proportion d'invariants non nuls rencontrés. Si tous les invariants pouvaient être égaux à zéro, cette somme serait égale à $\sum_H A_H$. Théoriquement il y aura toujours des invariants différents de 0 et comme nous l'avons vu, la valeur la plus probable de α_H peut être calculée à priori (α_{Hest}) d'où une valeur estimée $S_{\alpha est} = \sum_H \alpha_{Hest}$. Cependant la formule de la tangente tend à maximiser les valeurs de α_H car elle ne tient compte que de la valeur moyenne du sinus invariant. Surestimant, de ce fait, le nombre d'invariants nuls, elle tend à rendre tous les vecteurs entrant dans l'évaluation de α_H, quasiéquipollents et donc à augmenter le module du vecteur résultant (α_H). Il s'en suit, qu'en pratique la valeur S_α correspondant à la bonne solution sera comprise entre $\sum_H A_H$ et $S_{\alpha\ est}$.

- le facteur d'échelle absolu :

$$F_{obs.} = \frac{\sum_H \alpha_H - \sum_H \alpha_{H_{has}}}{\sum_H \alpha_{Hest} - \sum_H \alpha_{H_{has}}}$$

où $\sum_H \alpha_{Hhas}$ représente la somme obtenue, quand toutes les réflexions prennent une phase quelconque, au hasard. Elle est égale à $\sum_H (\sum_K A^2_{H,K})^{1/2}$ (23). F_{obs} est égal à zéro si les valeurs des phases sont distribuées au hasard, et voisin de 1 si les phases ont leur valeur réelle ($\sum_H \alpha_H = \sum_H \alpha_{Hest}$). Du fait du rôle maximisant de la formule de la tangente, la solution correcte aura en général un facteur F_{obs} supérieur à 1. Si une solution incorrecte peut également donner un facteur F_{obs} de cet ordre de grandeur, il est très peu probable qu'une solution ayant un facteur F_{obs} inférieur à 0,9 soit correcte (*).

(*) *En réalité le facteur F_{obs} et la somme $\sum_H \alpha_H$ représentent exactement la même grandeur, car les autres termes qui rentrent dans le rapport, sont calculées, à priori, indépendamment de la solution étudiée.*

Notons également qu'un test du type $R = \frac{\sum_H |\alpha_H - \alpha_{Hest}|}{\sum_H \alpha_{Hest}}$ *n'apporterait pas plus de renseignements que le facteur de Karle, car α_H et $E_{Hcalc.}$ représentent des grandeurs quasiproportionnelles.*

- le test en Ψ_o de Cochran et Douglas (17) doit être minimum. La relation définissant le facteur de structure calculé $E_{Hcalc.} = k \sum_K E_K E_{H-K}$ reste vraie si H est une réflexion faible, voire non observée.

La fonction Ψ_o est la somme , $k \sum | \sum E_K E_{H-K} |$, des modules des facteurs de structures normalisés des réflexions H faibles, calculés à partir de l'ensemble des réflexions K et H-K dont la phase a été évaluée par la formule de la tangente; elle est donc généralement faible.

Mis à part le calcul des invariants de structure que nous verrons au chapitre suivant, cette fonction constitue le seul moyen de faire intervenir l'information apportée par les réflexions non observées. Elle est, de plus, indépendante de l'algorithme utilisé au cours de la génération de phases et, comme aucun affinement n'est fait sur les réflexions non observées, elle donne des résultats plus contrastés que les figures de mérite qui portent sur les phases et les facteurs de structure affinés. Pour toutes ces raisons, elle constitue la figure de mérite la plus efficace.

Bien souvent, les tests S_α et Ψ_o donnent des résultats contradictoires : ils sont maxima et minima en même temps. Dans le groupe $P\bar{1}$, par exemple, la solution qui donne la plus grande somme S_α correspond à des phases invariantes toutes nulles, tous les triples produits $E_H E_K E_{H-K}$ sont positifs et par conséquent la somme $\Psi_o : \sum \alpha_H$ sur les réflexions faibles est également maximale. Par contre une valeur S_α faible correspond en général à des phases distribuées au hasard ; pour une telle solution, les termes entrant dans le test de Cochran seront également distribués au hasard, la valeur Ψ_o sera minimale.

Il se peut donc que la bonne solution ne soit pas classée première, à la fois sur le test S_α et le test Φ_o. Cependant, si on étudie n solutions différentes et si l'une d'elles est $n_1^{ième}$ et $n_2^{ième}$ sur S_α et Ψ_o, elle ne pourra être considérée comme acceptable qu'à la condition que $n_1 + n_2 < n$.

L'étude de différences du type :

$$\frac{S_\alpha}{S_{\alpha max.}} - \frac{\Psi_o}{\Psi_{omax.}} \quad \text{ou} \quad \frac{S_\alpha}{S_{\alpha est}} - \frac{\Psi_o}{\Psi_{oest}}$$

constituent, à notre avis, une meilleure utilisation de ces 2 tests.

Finalement, le choix de la bonne solution dépend :

- de la comparaison des résultats donnés par les relations $\sum_1$ avec les phases obtenues en fin d'affinement.
- du facteur résiduel de Karle
- de la différence $\frac{S_\alpha}{S_{\alpha est}} - \frac{\Psi_o}{\Psi_{oest}}$

L'expérience nous a montré que si aucune solution ne se détache nettement des autres sur ces 2 figures de mérite, aucune d'elles ne

donnera de sections de densités correctes. Il y a alors une erreur au cours de la génération de phases :

- ou un invariant de structure différent nettement de 0 a été rencontré dans les toutes premières phases générées : dans ce cas, on pourrait utiliser le contrôle symbolique des phases.

- ou l'assignation de phases au symbole est erronée ; l'une des solutions symboliques utilisées est incorrecte.

III-4 APPLICATION A LA RESOLUTION DE STRUCTURES STEROIDIQUES

Oestrone Sublimé $P2_12_12_1$ (9)

Le programme de multisolution de P.Main et G.Germain (MULTAN) a été utilisé pour les 216 réflexions ayant un facteur E supérieureà 1,40. Nous avons accepté comme déterminées par la relation $\sum_1$, les réflexions pour lesquelles la probabilité $P(\sum_1)$est supérieure à 0,90, c'est à dire :

2 0 6	E = 2,00	$\Phi = 180°$	$P(\sum_1) = 0,9285$
2 0 8	E = 2,20	$\Phi = 0°$	$P(\sum_1) = 0,9220$

Le processus de convergence alors appliqué à l'ensemble des réflexions propose comme origine et énanthiomorphe :

1 0 8	E = 3,37	$\Phi = 0°$
1 0 3	E = 2,66	$\Phi = 90°$
0 1 3	E = 1,56	$\Phi = 90°$
7 3 0	E = 3,12	$\Phi = 90°$

Après examen des valeurs α_{Hest} prises par chaque réflexion au moment de son élimination dans le processus de convergence, il fut décidé de travailler avec un seuil d'acceptation $\alpha_{min} = 3,60$ et donc d'utiliser comme symboles les 3 réflexions pour lesquelles la valeur α_{est},au moment de leur élimination,est inférieureà α_{min}. En effet, sur ces réflexions, l'utilisation de la relation $\sum_2$ pourrait entrainer des erreurs non négligeables.

multisolution initiale

0 0 4	E = 3,08	m(0 ou π)	$\alpha_{est} = 0$	0 ou π
5 7 2	E = 2,05	a	$\alpha_{est} = 0$	$\pm\frac{\pi}{4}$ ou $\pm\frac{3\pi}{4}$
2 9 1	E = 2,12	b	$\alpha_{est} = 1,68$	$\pm\frac{\pi}{4}$ ou $\pm\frac{3\pi}{4}$

En appliquant la méthode de la multisolution sous sa forme initiale, il y aurait donc 32 solutions à étudier.

Pour réduire le nombre de solutions, nous avons utilisé la méthode d'addition symbolique. Il apparait très rapidement que la relation symbolique $m = \pi$ (cette relation étant d'ailleurs confirmée par la relation $\sum_1$ qui indique que $\Phi_{004} = \pi$ avec une probabilité $P\sum_1 = 0{,}8258$).

Deux autres indications furent ensuite obtenues $a = \pi$ et $b = 0$. Ces symboles représentant des phases quelconques, l'affinement par la formule de la tangente a donc porté sur 4 ensembles de départ différents :

$$a = \pi \pm \frac{\pi}{4} \qquad \text{et} \qquad b = \pm \frac{\pi}{4}$$

Par affinement, les solutions correspondant à une même valeur de b convergent vers un même point. On n'a donc que 2 ensembles de figures de mérite à étudier :

	$b = +\pi/4$	$b = -\pi/4$
R_{Karle}	37,84	34,95
S_α	6016,7	6179,9
Φ_o	331,9	263,6
F_{ABS}	0,8256	0,8582
$\Delta_\alpha = S_\alpha - S_{\alpha est}$ (voir IIème Partie)	-0,1224	+0,1052

La solution $b = -\pi/4$ est la meilleure sur l'ensemble des figures de mérite. Pour cette solution, les valeurs finales des symboles sont $a = 161°$ et $b = -44°$. (Si d'après cette étude la valeur $a = 180°$ aurait pu être utilisée sans inconvénient pour la suite du calcul, on peut se demander si la solution $b = 0$ aurait converger vers la valeur correcte $b = -\frac{\pi}{4}$. Nous pensons que cette détermination illustre parfaitement les avantages de la combinaison des deux méthodes :

→ d'une part la réduction du nombre de solutions à étudier par l'addition symbolique : 4 au lieu de 32.

→ d'autre part la possibilité de corriger une indication symbolique légèrement fausse par la multisolution .

Une carte de Fourier faite à partir des phases ainsi obtenues a permis de placer sans ambiguïté, 19 des 20 atomes de la molécule. Il y avait seulement deux pics parasites.

L'écart moyen entre les phases déterminées par méthodes directes et celles obtenues en fin d'affinement (R = 0,069) est égal à $\langle \Delta\Phi \rangle_H = 11°6$.

CHAPITRE IV

METHODE DES INVARIANTS DE STRUCTURE

Si l'addition symbolique et la multisolution conservent un lien de parenté étroit, car elles dérivent, plus ou moins, de la valeur moyenne nulle du sinus invariant, la méthode des invariants de structure développée par H.Hauptman représente une technique radicalement différente, où l'on se propose d'évaluer le cosinus invariant à partir des modules des facteurs de structure. Du point de vue théorique, elle semblerait supérieure aux autres méthodes de détermination, puisqu'elle permet d'accéder directement à la valeur exacte de la phase invariante $\Phi_H + \Phi_K + \Phi_{H-K}$; nous verrons au cours de ce chapitre, que les différentes approximations que l'on est obligé de faire, diminuent singulièrement l'efficacité de cette méthode.

Ici aussi, nous essaierons d'employer la notion de "symbole" et de voir si l'utilisation d'une phase invariante non nulle, supposée plus proche de la réalité, améliore la précision de la détermination.

IV-1 PRINCIPE DE LA METHODE (H.Hauptman) (27,26)

Si l'on considère 3 réflexions de l'espace réciproque h_1, h_2 et et h_3 telles que $h_1+h_2+h_3 = 0$ (*), la somme de leurs phases $\Phi_1 + \Phi_2 + \Phi_3$ ne dé-

---

(*) *Pour des raisons de commodités nous avons provisoirement changé de notations en posant :*

$$-H = h_1 \; ; \; K = h_2 \; ; \; H - K = h_3$$

d'où on a immédiatement :

$$h_1 + h_2 + h_3 = 0$$

pend pas du choix de l'origine mais uniquement de la structure : c'est un invariant qui peut être calculé à priori, à partir des modules des facteurs de structure.

Il est, par exemple, possible d'évaluer le cosinus invariant par une relation statistique, qui, si tous les atomes sont répartis au hasard, peut se mettre sous la forme : (29c).

$$|E_{h_1} E_{h_2} E_{h_3}| \cos(\Phi_1 + \Phi_2 + \Phi_3) = \frac{N^{3/2}}{2} \langle (|E_k|^2 - 1)(|E_{k+h_1}|^2 - 1)(|E_{k-h_3}|^2 - 1) \rangle$$

$$+ \frac{1}{\sqrt{N}} (|E_{h_1}|^2 + |E_{h_2}|^2 + |E_{h_3}|^2 - 2)$$

(dans le cas d'atomes non identiques on remplace $\frac{1}{\sqrt{N}}$ par $\sigma_3 / \sigma_2^{3/2}$).

En réalité les atomes appartiennent à des molécules organisées ; certains motifs d'atomes se répètent plusieurs fois dans la structure, entrainant sur la fonction de Patterson, le renforcement de certains pics. On appelle ces motifs "interactions induites" (*).

Un pic multiple de l'espace de Patterson, du à un recouvrement fortuit de plusieurs pics (interactions fortuites), va également influer sur la valeur moyenne du triple produit.

La présence de ces différents types d'interactions entrainent une augmentation des valeurs moyennes de $\langle (|E_K|^2 - 1)^p$; il faut donc modifier la formule utilisée en conséquence : (27).

$$|E_{h_1} E_{h_2} E_{h_3}| \cos(\Phi_1 + \Phi_2 + \Phi_3) = \frac{\sigma_1^2 - 3\sigma_1 \sigma_2 + 2\sigma_3}{\sigma_2^{3/2} \langle (|E_K|^2 - 1)^3 \rangle_K} \langle (|E_K|^2 - 1)(|E_{K+h_1}|^2 - 1)$$

$$(|E_{K-h_3}|^2 - 1) \rangle_K$$

$$+ \frac{\sigma_3}{\sigma_2^{3/2}} (|E_{h_1}|^2 + |E_{h_2}|^2 + |E_{h_3}|^2 - 2)$$

Cette formule doit encore être modifiée pour tenir compte du fait que les recouvrements dans l'espace de Patterson ne sont que partiels

---

(*) Pour les stéroïdes oestrogènes, les 4 cycles hydrocarbonés sont tous différents : le cycle A est benzénique, le cycle B cyclohéxènique, le cycle C cyclohéxanique, le cycle D pentagonal ; de ce fait il existe très peu d'interactions induites entre les différents atomes. Il en va tout autrement pour les dérivés de l'androstane où les cycles A, B et C sont tous cyclohexaniques et font apparaître un grand nombre d'interactions induites.

Il semblerait, pour ces composés que la proportion d'invariants négatifs augmente sensiblement : l'étude de telles structures à l'aide de la relation $\sum_2$ ne fournit souvent qu'une solution partielle correspondant à une déconvolution incomplète de la fonction de Patterson.

l'existence de pics distincts ou confondus pour les interactions fortuites dépend du pouvoir de résolution des facteurs de structure utilisés. Pour en tenir compte on remplace le terme multiplicatif devant le triple produit par un facteur d'échelle K que l'on détermine de façon semi-empirique, en ajustant la distribution expérimentale des cosinus invariants, à la distribution théorique, pour $|E_{h_2}|$ et $|E_{h_3}|$ fixés.

La probabilité élémentaire pour que la phase invariante prenne la valeur $\Phi_1 + \Phi_2 + \Phi_3$ a déjà été définie (II.3) :

$$P(\Phi_1+\Phi_2+\Phi_3) = C' \, e^{A_{h_1,h_2} \cos(\Phi_1+\Phi_2+\Phi_3)} \, d(\Phi_1+\Phi_2+\Phi_3)$$

Si l'on fait le changement de variable $x = \Phi_1 + \Phi_2 + \Phi_3$, il vient :

$$P(x/|E_{h_2}|, |E_{h_3}|) = C' \frac{e^{Ax}}{\sqrt{1-x^2}} \, dx$$

En égalant à 1, l'intégrale de cette fonction étendue de $x = -1$ à $x = +1$, on détermine C' égal à $\frac{1}{\pi I_0(A)}$.

La fonction de distribution théorique dépendant de A_{h_1,h_2}, on peut considérer que le facteur d'échelle K est aussi une fonction de A_{h_1,h_2} : Hauptman utilise trois types de fonction K = Cte , K = aA+b et K = alog A.

Maintenant on connait la valeur exacte (aux erreurs d'expérience près) de la somme $\Phi_H + \Phi_K + \Phi_{H-K}$. Les phases des facteurs de structures peuvent alors être déterminées de façon sûre, dès le départ, une fois l'origine fixée. Dans ce but Hauptman préconise pour la détermination des phases Φ_H à partir des invariants, un processus de moindres carrés : Φ_H sera déterminée à partir de l'ensemble des Φ_K et Φ_{H-K} connues, en minimisant l'expression :

$$\frac{\sum_K W_K \left[\cos(\Phi_H + \Phi_K + \Phi_{H-K}) - c_K\right]^2}{\sum_K W_K}$$

où c_K désigne le cosinus invariant déterminé précédemment .

$W_K = |E_H \, E_K \, E_{H-K}| \sqrt{n}$ est le poids attaché à cette détermination (n représente le nombre de termes entrant dans l'évaluation du triple produit).

On détermine alors les différentes phases Φ_H, de proche en proche, jusqu'à ce que l'on en obtienne un nombre suffisant pour entreprendre un affinement par la formule de la tangente.

IV-2 ETUDE EXPERIMENTALE

L'élaboration d'un programme calculant le cosinus invariant par la formule du triple produit et l'étude expérimentale détaillée de cette méthode a déjà fait l'objet d'une thèse de la part de notre groupe de travail (G.Comberton 15) . Nous voulons simplement ici rappeler certaines particularités à partir desquelles nous essaierons de mettre au point une technique d'utilisation.

L'évaluation du triple produit constituerait le mode de détermination de structures le plus efficace, si la valeur obtenue par cette formule ne s'écartait pas trop de la valeur exacte. Nos résultats expérimentaux, comme ceux obtenus par Hauptman (26), montrent qu'il n'en est malheureusement pas ainsi, trop souvent la valeur calculée diffère notablement de la valeur exacte de la phase invariante.

On peut, d'ailleurs, comprendre facilement qu'il est impossible d'obtenir une évaluation correcte de la phase invariante par la formule du triple produit, car cette relation n'est rigoureuse que si les recouvrements dans l'espace de Patterson son exacts, et si la sommation porte sur tous les vecteurs de l'espace réciproque. Cependant il semble que l'influence de la limitation de la sphère sur le résultat est moins néfaste.

Néanmoins, si cette détermination reste imparfaite, il est possible de l'améliorer empiriquement. L'étude expérimentale faite par G. Comberton a montré en effet que la répartition des erreurs ne s'effectuait pas au hasard. Les résultats obtenus à partir des données de l'oestrone sublimé (dont nous avons vu le procédé de détermination au chapitre précédent), permettent de penser que les phases invariantes calculées proche de 0° ou de 180° sont vraisemblablement exactes, alors qu'une phase calculée voisine de 90°, a pratiquement autant de chance d'être bien déterminée que d'être entachée d'une erreur importante (qui peut même être égale à 90°).

Comme l'illustre le tableau ci-dessous les indications obtenues pour les phases extérieures à l'intervalle (60°, 135°) sont en moyenne plus proche de la réalité, que la valeur nulle utilisée dans la formule de la tangente et la relation $\sum_2$. Par contre à l'intérieur de cet intervalle, la marge d'erreurs est au contraire très élevée.

Valeurs de $A_{H,K}$ entre parenthèses : nombre d'invariants considérés	Ecart moyen du cosinus invariant calculé à la valeur exacte		Ecart moyen du cosinus exact à la valeur + 1,0 prise dans la formule de la tangente (pour la même répartition que précédemment).	
	extérieur à (60°,135°)	intérieur à (60°,135°)	extérieur à (60°,135°)	intérieur à (60°,135°)
> 2.50 (54)	0,1885 (40)	0,6678 (14)	0,2047	0,2547
2.0 < <2.50 (128)	0,2507 (95)	0,7394 (33)	0,3166	0,4321
1.8 < <2.00 (93)	0,2449 (61)	0,8116 (32)	0,3280	0,4778
1.7 < <1.8 (54)	0,3000 (37)	0,8358 (17)	0,3514	0,4800

Cette étude expérimentale a également montré que les triples produits pour lesquels le nombre de "contributions" est anormalement faible, sont généralement entachés d'une erreur importante.

IV-3 QUELQUES REFLEXIONS SUR LE FACTEUR D'ECHELLE

Comme nous l'avons dit, le facteur d'échelle K de la formude triple produit se calcule en ajustant la distribution expérimentale des cosinus invariants à la distribution théorique. Quelles sont les conséquences des résultats expérimentaux décrits ci-dessus sur ce calcul ?

Pour calculer K, on considère qu'il existe une certaine proportion P_1 de cosinus invariants supérieurs à cos 60°. En identifiant le plus petit de ces cosinus à cette valeur on détermine une valeur approchée de K. Mais nous avons vu au paragraphe précédent qu'une partie des invariants calculés entre 60° et 135° étaient faux , leurs valeurs réelles étant soit supérieures à 135°, soit plus vraisemblablement inférieures à 60°. (Pour ces invariants l'écart moyen du cosinus invariant à +1,0 est nettement plus faible que l'écart en cosinus calculé). Par la suite quand on prend les NP_1 plus forts cosinus invariants (N nombre d'invariants considérés), on ne tient pas compte de ces valeurs mal calculées. Le terme servant au calcul du facteur d'échelle devrait correspondre à un cosinus invariant plus faible que cos 60°. Le facteur d'échelle est, de ce fait surévalué.

Soit p_1, p_2, p_3 les proportions théoriques d'invariants respectivement inférieur à 60°, compris entre 60° et 135°, et supérieurs à 135°, et p'_1 , p'_2 , p'_3 les valeurs expérimentales qui tiennent compte des déterminations erronées. On a, du fait des considérations précédentes, $p'_2 = \alpha\, p_2$ avec $\alpha > 1$; si on admet alors que les variations de p_1 et p_3 restent proportionnelles on aura :

$$\frac{p'_1}{p_1} = \frac{p'_3}{p_3} = \frac{1-p'_2}{1-p_2} = \frac{1-\alpha p_2}{1-p_2}$$

La véritable proportion à partir de laquelle on doit calculer le facteur d'échelle est donc :

$$p'_1 = p_1 \times \frac{1-\alpha p_2}{1-p_2}$$

La valeur de α n'est pas connue, mais d'après les résultats expérimentaux rapportés ci-dessus,on peut l'estimer égale à 1,5 ou 2.

IV-4 ESSAIS D'AMELIORATION DU CALCUL DES COSINUS INVARIANTS

Le temps de calcul des cosinus invariants est déjà considérable si on le compare au temps nécessité par l'addition symbolique : il faut compter environ 30 mn de 360-44 pour calculer 2000 invariants pour une structure comme l'oestrone III où l'addition symbolique fournit une solution en moins de 10 mn. Tout essai d'amélioration va entrainer une augmentation du temps de calcul. Cette augmentation sera justifiée si les valeurs obtenues sont nettement meilleures que celles utilisées dans la formule de la

tangente. Nous décrirons ci-dessous 3 procédés d'améliorations que nous avons classés par ordre croissant de préférence en fonction de ce critère.

IV-4-1 La formule MDKS de H.Hauptman

H.Hauptman a proposé, dans le cas de structure ayant un grand nombre d'atomes ou présentant un nombre important de superpositions dans l'espace de Patterson, d'utiliser une évaluation du cosinus invariant plus souple que celle du triple produit (27b).

Le cosinus invariant peut être calculé par une expression de la forme :

$$\cos(\Phi_{H_1} + \Phi_{H_2} + \Phi_{H_3}) = M\,(D - KS)$$

où

$$D = \frac{1}{m_1+m_2+m_3}\,(m_1 D_1 + m_2 D_2 + m_3 D_3)$$

avec $D_i = \left\langle \left| (|E_{H_i+K}|^2 - 1) \,/\, |E_K| > t,\ |E_{H_j+K}| > t \right| \right\rangle_K$

$i = 1, 2, 3$ et $j = 2, 3, 1$ respectivement

m_i est le nombre de terme entrant dans la moyenne D_i

t un seuil fixé à l'avance (par exemple 1 ou 1,5)

et

$$S = \frac{1}{n_1+n_2+n_3}\,(n_1 S_1 + n_2 S_2 + n_3 S_3)$$

avec $S_i = \langle (|E_{-H_i+K}|^2 - 1) \,/ |E_K| > t \rangle_K + \langle (|E_{-H_i+K}|^2 - 1) \,/ |E_{H_j+K}| > t \rangle_K$

n_i est le nombre de termes entrant dans S_i

M et K sont des constantes évaluées semi-empiriquement.

Cette formule a été employée avec succès par H.Hauptman pour déterminer les invariants de structure relatifs aux plans principaux hk0, h0l et 0kl de la structure oestradiol + urée ($P2_12_12_1$) pour lesquels les résultats donnés par la formule du triple produit n'étaient pas satisfaisants. Par contre pour d'autres structures, il semble que le triple produit soit plus utile.

Le tout serait de savoir dans quel cas utiliser le triple produit ou la formule MDKS ?

IV-4-2 Utilisation des relations entre invariants

L'utilisation des relations existant entre invariants devrait permettre d'améliorer les résultats fournis par le triple produit,

aussi bien que par la relation MDKS. Nous avons vu en effet qu'il était possible de définir plusieurs invariants par réflexions utilisée. Il existe donc des relations entre les invariants de structures, le cas le plus simple étant représenté par les quadripoles de Vaughan.

Il est également possible d'obtenir des relations entre invariants à partir du processus de convergence. En considérant le sens de convergence remontant, quand on introduit une réflexion définie par n invariants, il est possible de trouver n-1 relations avec les invariants précédemment utilisés. Les relations de symétrie liées au groupe spatial considéré permettent généralement de réduire ces relations à des formes groupant 4 invariants, comme les formes quadripolaires (voir tableau IV-4-2-1).

Pour les structures centrosymétriques, nous bénificions du fait qu'une relation invariante entre phase ne pouvait prendre que deux valeurs différentes (0, π). Par l'opération précédente, nous retrouvons de nouveau la notion de relation exacte pour une structure non centrée ; c'est cette fois-ci une relation entre phases invariantes :

$$\Phi_1 + \Phi_2 + \Phi_3 + \Phi_4 = 0 \text{ ou } \pi$$

On voit tout de suite l'intérêt de telles relations : les différentes valeurs Φ_i ont été calculées indépendamment les unes des autres, mais elles sont entachées d'une certaine erreur. L'utilisation des relations entre invariants doit permettre d'affiner les valeurs obtenues.

A propos des quadripoles de Vaughan, J.Karle (29c) avait déjà signalé une autre utilisation possible des relations entre invariants : elles permettent de lever l'indétermination de signe qui persiste sur la phase Φ_i qui n'est connue qu'à partir de son cosinus. En effet, il n'existe généralement qu'une seule combinaison de quatre signes pour laquelle la relation entre invariants est vérifiée.

Ce mode d'affinement, relativement onéreux au point de vue temps de calcul, ne doit cependant être réservé qu'aux structures n'ayant pu être déterminées par des moyens de calcul plus simple.

IV-4-3 Pondération semi-empirique

Bien que tous les résultats ne soient pas absolument satisfaisants, l'étude expérimentale a montré que le calcul des invariants par la formule du triple produit ou par la moyenne conditionnelle MDKS (*) pouvait être utilisé comme détecteur d'invariants de structure voisins de π.

(*) D'après H.Hauptman l'utilisation de l'un ou l'autre mode de calcul, liée au nombre de superpositions dans l'espace de Patterson (nous verrons dans la IIème Partie que d'autres phénomènes peuvent causer l'apparition d'invariants négatifs) peut être décidée, à priori, après considération des valeurs moyennes $\langle(|E|^2-1)^2\rangle_K$ et $\langle(|E|^2-1)^3\rangle_K$ sur les réflexions générales et sur celles des plans principaux.

Si ces moyennes sont de l'ordre des valeurs théoriques (respectivement 1.00 et 2.00) la structure doit pouvoir être résolue sans difficulté par le processus d'addition symbolique et la multisolution. Si elles sont doubles ou triples des valeurs théoriques on peut utiliser le triple produit Au delà il vaut mieux utiliser MDKS.

Tableau IV-4-2-1 : Détermination des relations entre invariants à partir des résultats de convergence (Androstènedione $P2_12_12_1$).

(On remonte ici de l'origine vers les dernières réflexions éliminées)

N°Code Phase	Indices		Résultats de convergence	Relations entre invariants
19	130		$\pm \pi/2$	
21	470	réflexions	0 ou π	
27	021	origine	0 ou π	
98	104		0 ou π	
26	111	$\Psi_{26} = \Psi_{19} - \Psi_{27} + \pi + \Phi_1$		Φ_i désigne la phase des invariants pris dans l'ordre où ils sont introduits.
76	213	$\Psi_{76} = \Psi_{98} - \Psi_{26} + \Phi_2$		
78	123	$\Psi_{78} = -\Psi_{27} + \pi + \Psi_{98} + \Phi_3$		
		$\Psi_{78} \pi \ \Psi_{19} - \Psi_{76} + \pi + \Phi_4$		$\rightarrow \Psi_{78} = \Psi_{19} - (\Psi_{98} - \Psi_{19} + \Psi_{27} + \Phi_1 - \Phi_2) + \Phi_4$

On compare les 2 déterminations de Ψ_{78} d'où

$$-2\Phi_{19} - 2\Phi_{98} + \pi = (\Phi_1 + \Phi_4) - (\Phi_2 + \Phi_3)$$

comme $2\Phi_{19} = \pi$ $\quad 2\Phi_{98} = 0$

$$(\Phi_1 + \Phi_4) - (\Phi_2 + \Phi_3) = 0 \qquad (1)$$

85	583	$\Psi_{85} = \Psi_{21} + \Psi_{76} + \Phi_5$		
40	581	$\Psi_{40} = -\Psi_{85} + \Psi_{98} + \Phi_6$		
		$= \Psi_{21} + \Psi_{26} + \Phi_7$	$\rightarrow$	$\Phi_2 + \Phi_5 + \Phi_7 - \Phi_6 = 0 \qquad (2)$
110	664	$\Psi_{110} = \Psi_{40} - \Psi_{78} + \pi + \Phi_9$		
		$= \Psi_{27} + \Psi_{85} + \pi + \Phi_8$	$\rightarrow$	$\Phi_1 - \Phi_4 - \Phi_5 + \Phi_7 - \Phi_8 + \Phi_9 = 0 \qquad (3)$
			(4)+(5) $\rightarrow$	$\Phi_5 + \Phi_8 - \Phi_{11} + \Phi_{12} = 0 \quad (3bis)$

Suite Tableau IV-4-2-1

100 214 $\Psi_{100} = -\Psi_{26}+\Psi_{78} + \Phi_{10}$
$= -\Psi_{110}-\Psi_{21}+ \pi+\Phi_{11}$ $\Phi_1-\Phi_2-\Phi_5-\Phi_8+\Phi_{11}+\Phi_1-\Phi_3-\Phi_{10} = 0$

(1) → $\Phi_1 - \Phi_4 - \Phi_5 - \Phi_8 - \Phi_{10} + \Phi_{11} = 0$

(3) → $\Phi_7 + \Phi_9 + \Phi_{10} - \Phi_{11} = 0$ (4)

Ψ_{100} + - Ψ_{27} - Ψ_{76} + Φ_{12} → $\Phi_1 - \Phi_4 - \Phi_{10} + \Phi_{12} = 0$ (5)

103 324 $\Psi_{103} = - \Psi_{100} - \Psi_{19} + \Phi_{13}$
$= \Psi_{26} + \Psi_{76} + \Phi_{14}$ → $\Phi_4 + \Phi_{10} - \Phi_{13} + \Phi_{14} = 0$ (6)

Exemple d'utilisation :

- Pour déterminer le signe de la phase invariante

$\Phi_1 = 0°$ $\Phi_2 = \pm 28°$ $\Phi_3 = \pm 50°$ $\Phi_4 = 0°$

la relation $\Phi_1 + \Phi_2 + \Phi_3 + \Phi_4 = 0$ impose donc que Φ_2 et Φ_3 soient de signes contraires. Il y aurait dans ce cas 2 solutions à étudier $\Phi_2 = + 28°$ et $\Phi_2 = - 28°$.

- Affinement d'une phase invariante :

On peut par exemple redéfinir Φ_1 à partir de (1) et (5)

$\Phi_1 = \Phi_2 + \Phi_3 - \Phi_4$ et $\Phi_1 = \Phi_4 + \Phi_{10} - \Phi_{12}$

On peut classer les différents invariants calculés en trois catégories :

- on identifie à 0° ceux qui sont calculés inférieurs à 60°. Dans ce cas, l'écart moyen à la valeur réelle du cosinus invariant est, pour une valeur $A_{H,K}$ donnée, plus faible que la valeur moyenne théorique $\frac{I_1(A)}{I_0(A)}$. On pourrait considérer que cet écart moyen représente la valeur théorique d'une classe d'invariants de valeur $A'_{H,K}$ avec

$$A'_{H,K} > A_{H,K}$$

Cette première catégorie d'invariants est donc identifiée à zéro avec un poids supérieur à la valeur $A_{H,K}$ généralement utilisée dans les études statistiques (formule de la tangente).

- on attribue un poids identique aux phases invariantes calculées supérieures à 135° et que l'on identifie à 180°.

- Pour les invariants calculés entre 60° et 135°, l'approximation de la formule de la tangente ($\Phi = 0°$) est meilleure que les indications des invariants. Cependant l'écart moyen de la valeur correspondante du cosinus à la valeur réelle est plus grand que l'écart théorique, on fait donc correspondre à la valeur $\Phi = 0°$ un poids $A''_{H,K}$ plus faible que $A_{H,K}$ et tel que

$$\frac{I_0(A'')}{I_1(A'')} = < \Delta \Phi >_{exp.}$$

Tenant compte des nouveaux poids attribués à chaque invariant, on pourra faire un calcul de α estimé et un processus de convergence,puis résoudre la structure par les techniques décrites aux chapitres précédents.-

IV-5 CALCUL DES PHASES ET SYMBOLES

Les différentes études faites jusqu'ici sur le calcul d'invariants(15,26)ont montré qu'en pratique l'erreur commise sur les cosinus invariants étant loin d'être négligeable. Dès lors on peut se demander si le temps de calcul élevé que nécessite l'affinement par moindres carrés proposé par Hauptman, se justifie actuellement : la précision qu'il apporte nous parait illusoire.

Nous lui préférerons le calcul direct de la phase résultante à partir d'une simple moyenne arithmétique, sous la forme :

$$\Phi_H = \frac{\sum_K W_K(|\Phi_K + \Phi_{H-K}| - \Theta_{H,K})}{\sum_K W_K}$$

où $\Theta_{H,K}$ représente la phase invariante.

Cette dernière valeur est calculée à partir du cosinus invariant c_K ; elle n'est par la suite déterminée qu'au signe près ($\pm \Theta_{H,K}$). Une fois l'énanthiomorphe choisi, s'il y a lieu, le signe de toutes les phases invariantes $\Theta_{H,K}$ est fixé (il n'est plus arbitraire), mais à priori on ne sait pas quelle en est la valeur.

On est alors obligé de donner successivement à chacun des invariants qui entrent dans le calcul d'une nouvelle phase Φ_H, les 2 valeurs possibles $\pm \Theta_{H,K}$. Pour une phase Φ_H déterminée par n invariants on obtient donc 2^n solutions possibles (beaucoup d'invariants étant nuls il y a en réalité moins de solutions). On sélectionnera alors pour Φ_H la détermination qui aura la plus petite variance :

$$\mathrm{Var}(\Phi_H) = \frac{\sum_K W_K \left| (\Phi_K + \Phi_{H-K}) - \Theta_{H,K} - \Phi_H \right|^2}{\sum_K W_K} = \frac{\sum_K W_K \left| (\Phi_K + \Phi_{H-K}) - \Theta_{H,K} \right|^2 - \Phi_H^2}{\sum_K W_K}$$

(Cette deuxième forme permettant d'évaluer simultanément dans le même calcul la moyenne Φ_H et sa variance).

Notons que, bien que cela n'apparaisse pas dans notre exposé du paragraphe IV-1, l'affinement par moindres carrés présente les mêmes difficultés, la résolution de l'équation correspondante faisant apparaitre des termes sinus $\Theta_{H,K}$ où l'on retrouve la même indétermination.

Cependant dans le calcul de la moyenne arithmétique intervient un 2ème type d'indétermination, du au fait qu'une phase n'est définie qu'à $2k\pi$ près; il est nécessaire pour chaque réflexion H, de faire encore 2 évaluations successives en utilisant pour les phases $\Phi_K + \Phi_{H-K} - \Theta_{H,K}$, deux définitions différentes correspondant respectivement aux intervalles $[-\pi, +\pi]$ et $[0, 2\pi]$.

Si, enfin, pour une même réflexion, on obtient deux déterminations Φ_H et Φ'_H associés à des variances peu différentes, on sera obligé d'utiliser une multisolution. Chaque solution sera associée à une figure de mérite qui sera la somme des variances $\sum_H \mathrm{Var}(\Phi_H)$ et l'on ne retiendra que les solutions dont la somme n'excède pas de plus de 10% la somme minimale.

Ce procédé de calcul des phases Φ_H présente le gros avantage de permettre l'utilisation de symboles dans les premiers tours de la détermination. Au lieu d'obtenir une détermination a ou π+a comme dans l'addition symbolique, on aura des valeurs "plus fines" du type :

$$a + \frac{\sum_K W_K \left((\Phi_K + \Phi_{H-K}) - \Theta_{H,K} \right)}{\sum_K W_K}$$

Cependant on éprouve davantage de difficultés dès qu'apparait pour une réflexion une détermination secondaire, car on ne peut utiliser la variance pour définir les signes des phases invariantes entrant dans

cette deuxième détermination. Pour pouvoir lever cette indétermination il faudrait connaitre la relation qui existe entre les 2 déterminations symboliques.

L'utilisation de symbole avec les invariants de structure ne deviendra efficace que si l'on peut déterminer le signe de la phase invariante par un autre moyen.

IV-6 APPLICATION A LA DETERMINATION DE L'ANDROSTENEDIONE

Nous avons appliqué la technique décrite précédemment à la détermination de la structure de l'androsténedione ($P2_12_12_1$, Z = 4) C_{18} O_2 H_{22} ; l'origine et l'énanthiomorphe ont été fixés par le processus de convergence :

130	$\pi/2$
470	0
021	0
104	0

La deuxième réflexion générée, 213, n'est définie que par un seul invariant (voir tableau du paragraphe IV-4-2) : il a été nécessaire d'étudier deux solutions différentes correspondant respectivement aux déterminations positives et négatives de la phase invariante.

Nous n'avons déterminé par cette technique, qu'une quarantaine de phases car les erreurs commises dans l'évaluation des invariants sont plus importantes dès l'instant où l'on fait appel à des invariants ayant une valeur $A_{H,K}$ faible.

A partir de ce stade, nous avons utilisé l'affinement par la formule de la tangente pour chacune des 2 solutions possibles. Les figures de mérite définies au chapitre précédent ont permis de sélectionner sans ambiguïtés la bonne solution. Les sections de densité obtenues alors permettent de placer les 21 atomes de la molécule et il n'y a aucun pic érroné.

L'écart moyen entre les phases déterminées par cette méthode et celles obtenues en fin d'affinement (R = 0, 04_1, (13)) est égal à $< \Delta \Phi >_H = 9°8$.

IV-7 CONCLUSION

A notre avis, le calcul des phases invariantes à partir des modules des facteurs de structure devrait s'avérer vraisemblablement supérieurs aux autres méthodes directes de détermination : c'est en effet la seule technique où l'on tienne compte de l'ensemble des observations, ainsi (mais toutefois à un degré moindre) que de la nature de la structure cristalline. On peut le considérer comme l'équivalent dans l'espace réciproque, de la fonction de Patterson définie sur l'espace réel.

Nous pensons que dans les prochaines années un effort important devrait être entrepris pour améliorer ce calcul, sans toutefois trop

augmenter son coût en heures d'ordinateur. Dans cet ordre d'idée, ce sont les techniques basées sur l'évaluation d'expressions analogues au triple produit qui semblent devoir donner les meilleurs résultats.

CONCLUSION DE LA PREMIERE PARTIE

Nous avons essayé, au cours de cette partie, d'exposer le plus simplement possible, les différentes méthodes directes qui sont habituellement proposées au cristallographe pour déterminer des structures. Nous avons donné pour chacune de ces méthodes des exemples d'application sur des structures relativement complexes (c'est à dire comptant une vingtaine d'atomes; nous avons vu qu'une structure correspondant seulement à une douzaine d'atomes ne poserait probablement aucun problème par l'addition symbolique classique) mais ne posant que des problèmes relativement simples puisque correspondant à des groupes pseudo centrés. Nous avons, de plus, proposé diverses améliorations qui devraient permettre d'aborder des problèmes plus compliqués (comme le montrera d'ailleurs la suite de ce travail). Cependant, malgré ces améliorations, il persiste un grand nombre de cas où la solution ne peut être obtenue de façon automatique. A la rigueur, quelque astuce permettra de résoudre une structure particulière, mais la stratégie alors définie ne sera généralement pas reproductible. Pour pouvoir progresser dans la résolution des structures, il nous semble nécessaire de se rappeler que l'analyse statistique, proposée par les méthodes directes, correspond à un problème physique, donc concret. C'est l'étude de ce nouvel aspect que nous nous proposons d'étudier maintenant.

D E U X I E M E P A R T I E

NOUVELLES CONCEPTIONS, NOUVELLES METHODES

INTRODUCTION

Dans de nombreux cas, les procédés d'analyse classiques ne fournissent pas une solution physiquement exploitable. Ces structures, que l'on qualifierait trop hâtivement d'inextricables, nécessitent une étude particulière. Peut être après avoir analysé les raisons de l'échec des méthodes classiques, il sera possible de définir de nouvelles règles de stratégie.

Les méthodes directes sont résolument tournées vers l'espace réciproque et ne regardent l'ensemble des données que par l'intermédiaire des mathématiques. A la fin du calcul, quand on a obtenu les sections de Fourier, on se rappelle que ces données sont intimement liées à une structure cristalline.

Mais la génération des phases, ce processus mathématiques que l'on confie généralement à un ordinateur, n'aurait elle pas une interprétation physique ? Ne serait-il pas possible de redéfinir toutes les notions que nous avons rencontrées précédemment, invariants de structure, choix de l'origine,... en termes physiques c'est à dire en termes de molécules et d'atomes ?

CHAPITRE I

INTERPRETATION PHYSIQUE DE LA GENERATION DE PHASES PAR METHODES DIRECTES ET APPLICATIONS

I-1 INVARIANTS DE STRUCTURE

(Pour simplifier l'exposé nous ferons un raisonnement pour une structure centrosymétrique, les résultats obtenus s'appliquent sans difficultés aux structures non centrées).

Considérons un facteur de structure normalisé $E_H(hkl)$ il peut se mettre sous la forme :

$$E_{hkl} = \frac{2}{\sqrt{\sigma_2}} \sum_{j=1}^{N} Z_j \cos 2\pi(hx_j+ky_j+lz_j)$$

Un atome j aura une contribution maximale si $hx_j+ky_j+lz_j = \frac{n}{2}$ (n entier). Si la structure étudiée n'est pas trop importante, une réflexion H pourra n'être influencée que par un seul atome (c'est le cas par exemple où j est un atome lourd).

Soit E_H et $E_{H'}$, deux réflexions fortement influencées par l'atome j :

$$hx_j + ky_j + \ell z_j = \frac{n}{2}$$

$$\text{et } h'x_j + k'y_j + \ell' z_j = \frac{n'}{2}$$

A ce moment, la contribution de l'atome j à la réflexion H-H' est également maximale car :

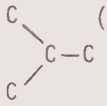
$$(h-h')\, x_j + (k-k')\, y_j + (\ell - \ell')\, z_j = \frac{n-n'}{2}$$

Si cette contribution est prépondérante, le signe de cette réflexion sera égal à $(-1)^{n-n'}$. Les signes de E_H et de $E_{H'}$ étant respectivement égaux à $(-1)^n$ et $(-1)^{n'}$, le signe de l'invariant de structure $\bar{H}$, H' et H-H' sera :

$$(-1)^n \times (-1)^{n'} \times (-1)^{n-n'} = (-1)^{2n} = +1$$

Dans ce cas, l'invariant de structure $\bar{H}$, H', H-H' positif est, en quelque sorte, "l'image" de l'atome j prépondérant. En réalité, une même réflexion n'est pas influencée par un seul atome (surtout dans le cas où il n'y a que des atomes légers). Si la contribution d'un atome est maximale pour une réflexion H, les contributions des atomes voisins (c'est à dire qui lui sont liés par covalence) seront également importantes et de même sens (*). Par contre on peut admettre en première approximation que la contribution des atomes éloignés est nulle,en moyenne.

Dans une structure organique, les différentes réflexions seront déterminées par des groupes d'atomes présentant une forte densité électronique :

- le cycle à 6 carbones (assimilable à 6 x Z_C électrons réunis en son centre).
- l'ensemble tétragonal (5 x Z_C électrons réunis sur le carbone central).
- l'ensemble trigonal  (4 x Z_C électrons réunis sur le carbone central).

(*) *Ceci n'est vrai que pour les réflexions de la sphère de* ***résolution*** *de 1,5 Å .*

En réalité, un invariant de structure peut correspondre à deux groupes adjacents, ayant en commun un ou plusieurs atomes. Pour une structure non centrée, cet invariant de structure "image" d'atomes voisins n'a pas en général une phase nulle; cette valeur ne se rencontrerait que si tout le nuage électronique correspondant était réuni au centre de gravité. La valeur obtenue s'écarte d'autant plus de la nullité que l'on s'éloigne de ce cas idéal. On peut être assuré qu'un invariant, mettant en jeu des atomes non liés par covalence, aura une phase qui s'éloigne passablement de zéro.

Les considérations précédentes permettent de comprendre qu'une phase invariante nulle soit la plus probable puisqu'elle correspond en général à un même groupe d'atomes. L'apparition d'une phase invariante voisine de π n'est qu'un évènement fortuit qui nécessite l'existence de 3 réflexions H, H' , H-H' dépendant d'au moins deux groupes d'atomes différents.

Il est cependant possible d'imaginer une telle éventualité : considérons par exemple une structure cristalline $P2_12_12_1$ constituée d'une molécule principale M allongée parallèlement à Ox et située en $y = (2k+1)\pi/8$ et d'une inclusion moins volumineuse I située en $y = 2k\pi/8$(par exemple la structure oestradiol + alc.propylique (voir proj.en annexe)).

Dans ce cas, la réflexion 0 4 $0_{(H)}$ est forte et sa phase égale à π est imposée par la molécule M ; cependant la contribution de l'inclusion est l'inverse ($\Phi = 0$ ou $S(E_H)_I = +$).

Soit H' et H-H', 2 réflexions fortement influencées par l'inclusion mais qui ne dépendent pratiquement pas de la molécule principale. Si on raisonne, comme précédemment, sur l'inclusion I, on peut écrire en ne considérant que la contribution de I à la réflexion H :

$$s\,(E_{H'}) \times s\,(E_{H-H'}) \times s\,(E_H)_I = +$$

Mais en réalité $s(E_H) = - s(E_H)_I$. Par la suite l'invariant H, H' , H-H' est négatif. Ce phénomène très ennuyeux dans les déterminations de structures apparait dès l'instant où l'inclusion I devient suffisamment importante (c'est à dire dès qu'elle groupe plus de 3 carbones), pour influencer des réflexions H et H-H'. Il est d'autant plus génant qu'il se rapporte alors à des réflexions fortes donnant de nombreux invariants de structure.

I-2 CHOIX DE L'ORIGINE ET GENERATION DE PHASES

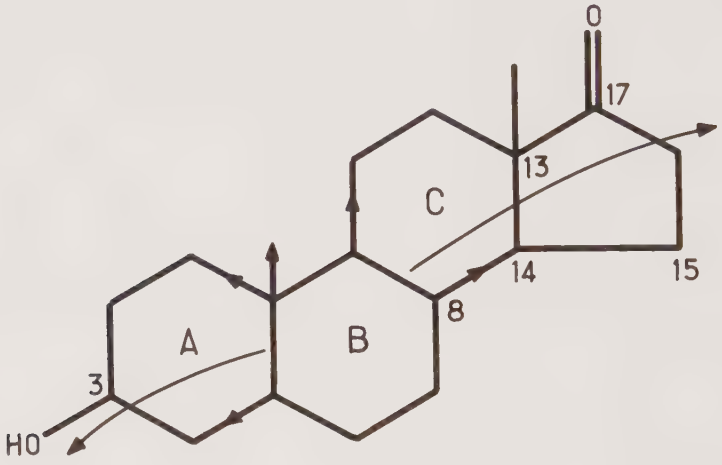
Pour déterminer les phases d'une structure par méthodes directes, il faut d'abord faire choix de réflexions "origine". On prend de préférence des réflexions fortes, donnant de nombreuses interactions (ou invariants de structure). A quoi correspondent-elles dans l'espace réel ?

- Etant fortes, elles correspondent à des zones où la densité électronique moyennes est élevée, c'est à dire où le nombre d'atomes lourds liés par liaisons covalente est important. Nous désignerons ce groupe d'atomes sous le nom de groupe "origine".

Donnant de nombreuses interactions, elles sont l'image d'un groupe qui est lié par covalence à un grand nombre d'atomes ou groupe d'atomes.

Dans le cas d'un stéroïde, les groupes les plus denses sont constitués par les différents cycles à 6 carbones A, B ou C. Mais le cycle A possède moins de "liaisons" que les cycles centraux. Le groupe "origine" sera donc l'un de ces cycles.

La génération des phases se fait alors de proche en proche, en utilisant les invariants de structure qui font intervenir les réflexions origine. Les premières réflexions, déterminées de cette manière, seront les "images" d'atomes directement liés au groupe "origine". Puis, à mesure que la génération avancera, l'utilisation d'invariants comprenant les nouvelles réflexions déterminées, permettra peu à peu de progresser vers les extrémités de la molécule.



Pour illustrer ces quelques notions, nous avons étudié les renseignements obtenus dans le cas la structure de l'oestrone I à différents stades de la génération de phases. Au départ les réflexions origine correspondent uniquement au cycle B. Peu à peu apparaissent les atomes périphériques, en commençant par les atomes multi liés C_3, C_{13}, C_{14}, C_{17}.

Comme les erreurs ont tendance à s'accumuler au cours de la génération, les dernières phases déterminées sont toujours les moins précises. On comprend dès lors que dans les différentes structures de stéroïde que nous avons étudiées de cette manière, les atomes C_{15} et C_{16}, situés loin des atomes "origine", soient toujours les plus mal déterminés et n'apparaissent généralement pas sur les sections de densité (cf Ière Partie II-3 et III-4).

La détermination d'une molécule stéroïdique, comme nous venons de la décrire, n'est pas sans rappeler la résolution de la structure de l'androstérone par DF.High et J.Kraut,à l'aide des fonctions minimum : "la première fonction permit d'identifier les 10 atomes des cycles B et C qui furent alors utilisés comme sites atomiques pour un nouveau calcul de superpositions. Puis apparurent le cycle A et l'hydroxyle O_3 ainsi que les 2 groupements C_{18} et C_{19}. Une dernière fonction révéla les atomes C_{17} et O_{17} mais les 2 derniers atomes (C_{15} et C_{16}) ne purent être localisés de cette manière" ; Quelle remarquable parallélisme existe entre méthodes directes et indirectes !

I-3 INFLUENCE DE LA FONCTION MAXIMISANTE DE LA FORMULE DE LA TANGENTE

Utiliser la formule de la tangente qui ne tient compte que de la valeur moyenne nulle du sinus invariant mais néglige complètement la valeur moyenne des cosinus invariants (voir lère Partie I-3), revient en quelque sorte à identifier implicitement les phases invariantes à zéro. (Ceci sera d'autant plus vrai que le nombre de termes utilisés dans chacunes des formules tangente sera faible).

Or nous avons vu précédemment qu'une phase invariante identiquement nulle ne se rencontrait que pour un groupe d'atomes réduit à son barycentre. La formule de la tangente aura donc tendance à favoriser les centres de gravité possibles, c'est à dire les atomes multiliés au détriment des autres.

Les différentes études de composés stéroïdiques, exposées dans la Ière partie sont en bon accord avec ce raisonnement.

- dans les structures de l'oestrone I et III nous avons déterminé respectivement 142 et 175 phases pour 20 atomes à placer; l'erreur commise dans cette évaluation est de l'ordre de 12° dans les 2 cas. Pour ces 2 structures, il est impossible de placer l'atome C_{15} et des pics parasites apparaissent au centre des cycles A, B, C (centre de gravité possible du groupe cyclique de 6 carbones).

- au contraire, dans la détermination de l'androsténedione, nous n'avons considéré que 118 réflexions alors qu'il y avait 21 atomes à placer, mais, du fait de l'utilisation préalable du calcul des invariants, nous avons pu bénéficier d'une détermination plus précise, où l'erreur moyenne commise n'est plus égale qu'à 9°8. Dans ce cas, non seulement tous les atomes apparaissent sur les sections, mais il n'y a plus aucun pic parasite. L'apparition d'un pic parasite au centre d'un cycle à 6 carbones ne semble donc pas du au fait que l'on n'utilise que les facteurs de structure les plus élevés (comme un raisonnement simple sur la molécule de benzène permet de l'imaginer). Pour

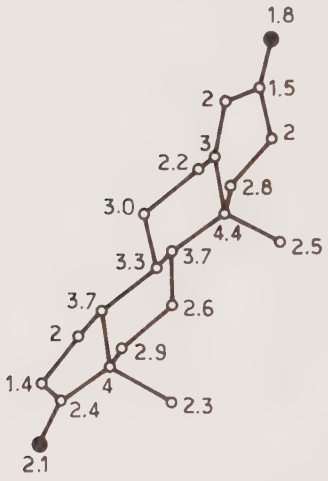
des molécules de cette compléxité elle est plutôt la conséquence d'une détermination erronée d'un certain nombre de phases.

Malgré la bonne détermination obtenue pour l'androsténedione cette structure illustre également le rôle néfaste joué par la formule de la tangente; on peut voir sur la figure ci-dessous :

- que les atomes C_{10} et C_{13} centre de gravité de motifs tétragonaux ont un poids respectif de 4,4 et 4.

- que les atomes C_8, C_9 et C_{14} centre de motifs trigonaux ont un poids compris entre 3 et 3,8.

alors que tous les autres atomes ont un poids inférieur à 3.



CHAPITRE II

NOTIONS D'ILES MOLECULAIRES

Les différents invariants de structure (ou relations $\sum_2$) relient les réflexions de l'espace réciproque les unes aux autres et permettent de fixer leurs phases relatives. Nous dirons qu'ils définissent un ensemble de corrélations parmi les réflexions de l'espace réciproque.

Si, comme on le suppose généralement dans les méthodes directes, les atomes sont répartis au hasard, il n'existe qu'un seul ensemble de corrélation : toutes les phases des réflexions de l'espace réciproque peuvent être générées, sans ambiguïtés, à partir des réflexions "origine".

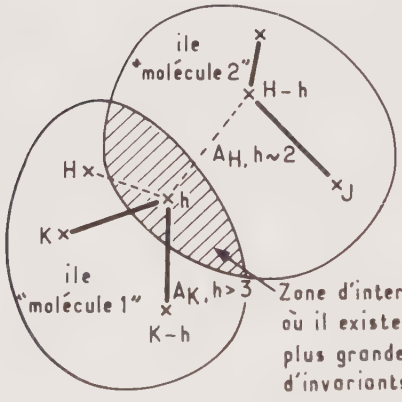
En pratique il n'en est jamais ainsi, car les atomes appartiennent à des molécules organisées, dans lesquelles leur position, les uns par rapport aux autres, sont imposées par des liaisons de covalence orientées; il n'est alors pas obligatoire que toutes les réflexions de l'espace soient liées entre elles : il peut apparaître des sous-ensembles de réflexions "fortement liées", ces sous-ensembles étant au contraire faiblement liés les uns aux autres, voire même complètement disjoints. Généralement, on désigne ces sous-ensembles de corrélation sous le nom d'iles.

On retrouvera cependant la notion d'atomes répartis au hasard, dès l'instant où, le nombre d'atomes dans la maille étant élevé, les liaisons

de Van der Waals deviennent du même ordre de grandeur que les liaisons de covalence, par rapport aux dimensions de la maille. Dans ce chapitre nous nous interesserons uniquement au problème posé par les petites molécules (au stéroïde par exemple) pour lesquelles la différence qui existe entre liaison de covalence et liaison de Van der Waals définit des relations différentes entre les atomes de l'espace réel.

II-1 EXISTENCE D'ILES MOLECULAIRES

Considérons une structure cristalline contenant plusieurs molécules (par exemple "molécule 1" , "molécule 2", ...) dans l'unité asymétrique, et supposons que les atomes "origine" appartiennent à une même molécule. A l'aide des invariants les plus "forts" on va "déterminer" de proche en proche les atomes qui leur sont liés par covalence, comme nous venons de le voir au chapitre précédent. Les invariants de structure "images" d'un groupe d'atomes liés par covalence ayant une phase proche de zéro, la valeur relative des phases ainsi déterminées sera correcte. Nous obtiendrons ainsi un système cohérent K, h, K-h (voir figure ci-dessous) et en utilisant uniquement les relations invariantes"fortes"(par exemple avec un $A_{K,h} > 3,0$) nous aurons finalement une connaissance parfaite de la molécule sur laquelle a été définie l'origine, mais aucun renseignement sur les autres molécules de l'unité asymétrique. Nous avons déterminé les phases de l'ile "molécule 1".



A ce stade, pour "atteindre" les atomes de la molécule 2, il faudra obligatoirement "passer" par des liaisons intermoléculaires : dans l'espace réel, entre les 2 molécules de la maille asymétrique, il y a une zone de faible densité électronique, correspondant aux atomes d'hydrogène, dont "l'image" dans l'espace réciproque sera un ensemble de réflexions faibles ou tout au moins faiblement liées par des relations invariantes. Ainsi la réflexion H-h qui correspond à la "molécule 2" appartient à une ile réciproque différente.

Or nous avons vu précédemment qu'un invariant nul correspondait généralement à des points de l'espace réel proche les uns des autres; par la suite, il y a au contraire de fortes chances pour que les invariants de structure permettant de passer d'une ile à l'autre, et qui correspondent à des atomes éloignés les uns des autres, différent notablement de zéro : la valeur obtenue pour la phase Φ_{H-h} sera donc entachée d'une certaine erreur (plus de 60° par exemple).

Mais ensuite, n'importe quelle phase Φ_j de l'ile "molécule 2" sera déterminée de façon correcte par rapport à cette phase Φ_{H-h} . L'ile "molécule 2" formera également un système cohérent et sera réellement représentative de la molécule 2 isolée, mais la translation de phase imposée par

Φ_{H-h} à toutes les phases de l'ile, correspond à une translation $\vec{\Delta}_{r0}$ de cette molécule.

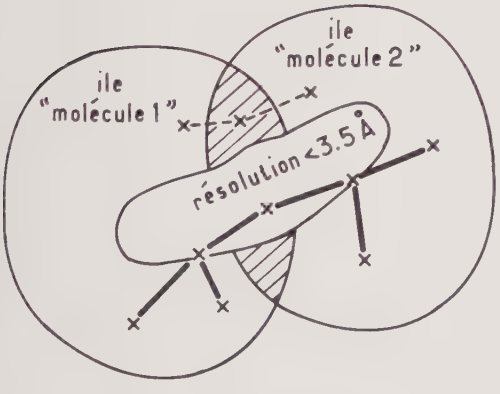
En pratique le résultat final est plus compliqué car le passage d'une ile à l'autre ne se fait pas par un seul invariant : plusieurs invariants de l'intersection interviennent successivement introduisant des translations de phases qui ne sont pas toutes de même sens. Sur les sections de densité on n'obtiendra alors qu'une portion translatée de la molécule 2, voire même une molécule complètement disloquée sous forme de groupes de 3 ou 4 atomes, translatés dans des directions différentes.(Voir plus loin III-2-4).

Pour que l'on puisse obtenir une détermination satisfaisante des phases de l'ile "molécule 2" il faut faire intervenir en même temps tous les invariants relatifs à l'intersection : on aurait à ce moment là l'équivalent d'un système d'équations à une inconnue (c'est à dire la véritable translation $\vec{\Delta}r_0$ qui permet de passer d'une molécule à l'autre).

On entrevoit alors le sens profond contenu dans la notion de symbole : alors qu'à l'intérieur d'une même ile, il n'est théoriquement pas besoin d'introduire de symboles pour générer l'ensemble des réflexions de l'espace à l'aide des relations $\sum_2$ fortes, on est obligé d'utiliser des symboles dès que l'on a affaire à des iles moléculaires : en repoussant au maximum la détermination du symbole "caractéristique" de l'ile "molécule 2", on doit réussir finalement à utiliser en même temps tous les invariants relatifs à l'intersection. Les diverses molécules de l'unité asymétrique sont alors situées correctement les unes par rapport aux autres.

II-2 INFLUENCE DU POUVOIR DE RESOLUTION

La notion d'iles moléculaires est liée à la présence d'une "cassure" à l'intérieur de l'unité asymétrique c'est à dire que l'on peut faire une différence entre la liaison de covalence et celle de Van der Waals. Encore faut-il pouvoir les distinguer l'une de l'autre.



On est ainsi amené à faire intervenir la notion de pouvoir de résolution et à modifier quelque peu le schéma des iles donné précédemment : les réflexions situées à l'intérieur de la sphère de résolution inférieure à 3,5 Å ne participent pas à l'existence d'iles moléculaires. Elles forment entre elles un système cohérent qui peut être déterminé en n'utilisant que des relations $\sum_2$ fortes.

En pratique, pour de petites molécules, la sphère de résolution 3,5 Å ne contient que peu de réflexions et de ce fait ne fournit qu'un nombre limité de relations $\sum_2$. On ne peut générer les phases des réflexions de la sphère de basse résolution de la même façon que pour les autres réflexions.

Cependant on peut utiliser cette sphère de basse résolution, tout en travaillant sur toutes les réflexions de l'espace : les invariants de la zone d'intersection correspondent à des valeurs A faibles; au contraire, ceux de la sphère de basse résolution correspondent à des valeurs de A fortes. Si au cours de la génération de phases on n'utilise que les relations invariants les plus fortes (par exemple A > 3 ou 4) on est assuré que le passage d'une ile à l'autre se fera essentiellement par la sphère de basse résolution plutôt que par la zone d'intersection. On peut espérer de cette façon obtenir finalement une meilleure détermination.

CHAPITRE III

MODE DE RESOLUTION DES SYSTEMES A ILES MOLECULAIRES

Deux cas sont à envisager, suivant que l'on a déjà placé une des molécules du système (par exemple parce que cette molécule comporte un atome lourd et a pu être localisée à partir de la fonction de Patterson, ou parce qu'elle a été déterminée lors d'une génération précédente qui n'a fourni qu'une solution partielle) ou non.

III-1 CAS OU UNE DES MOLECULES EST CONNUE

III-1-1 Technique de Résolution Générale

Dans ce cas commençons par une mise en garde : il est complètement illusoire de croire qu'en utilisant les phases déterminées par cette première molécule et en essayant d'en générer d'autres par la formule de la tangente,on arrivera à la solution correcte. Ce qui est parfaitement justifié quand on travaille avec une solution partielle d'un problème à une seule ile moléculaire devient catastrophique dès l'instant où il y en a plusieurs.

Il y a cependant une manière élégante de résoudre le problème, car,connaissant une des molécules de l'espace réel,on connait parfaitement l'une des iles de l'espace réciproque. Les réflexions qui ne sont pas calculées avec un facteur de structure Fc suffisant (40 % de Fo par exemple)

appartiennent à la deuxième ile. Celles pour lesquelles la valeur des Fc est compris entre 40 et 70 % de Fo peuvent être considérées comme appartenant à la zone d'intersection. Il suffit alors de déterminer par le processus de convergence, la réflexion de cette ile la plus judicieuse pour servir de symbole, puis de générer les différentes réflexions de l'ile en fonction de ce symbole. On peut alors aisément lever l'indétermination restante en calculant ce symbole à l'aide de la zone d'intersection.

III-1-2 Etude de structures à atomes lourds

La résolution de structure de bromodérivés (ou composés analogues,où un atome lourd est lié par covalence à la molécule étudiée) est relativement aiséedans le domaine centrosymétrique. Elle devient plus délicate pour les groupes non centrés,où les transformées de Fourier,faites avec les phases de l'atome lourd,font apparaitre des dédoublements des pics atomiques qui correspondent à une déconvolution imparfaite de la fonction de Patterson.

Le problème s'aggrave dès l'instant où l'on a affaire à un bromhydrate où l'atome de brome n'est pas lié par covalence à la molécule principale. Dans ce cas, les techniques de l'atome lourd deviennent nettement moins efficaces.

Etant donné qu'il existe dans l'unité asymétrique deux ensembles moléculaires différents (l'atome lourd d'une part, la molécule étudiée de l'autre), on aura dans l'espace réciproque deux iles différentes. La zone d'intersection de ces iles ne comportent que des invariants faibles correspondant à des atomes voisins dans les deux ensembles moléculaires; nous pouvons dire, en première approximation, qu'ils sont "l'image" des contacts de Van der Waals existant entre ces deux molécules.

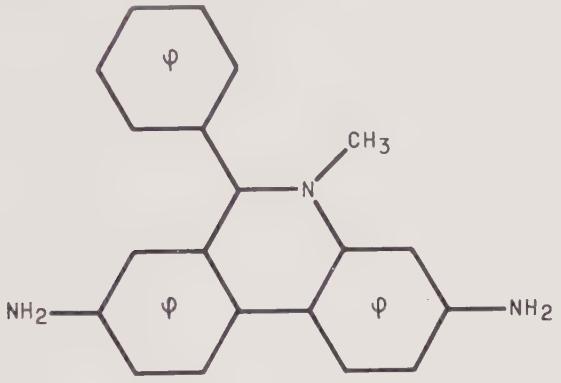
Quand l'une des iles est "l'image" d'un atome lourd isolé, les électrons qui, dans le cas général, étaient répartis sur une certaine surface, forment ici un ensemble ponctuel, et,alors qu'il existe de nombreux contacts entre 2 molécules relativement volumineuses, nous n'en rencontrons qu'un petit nombre entre la molécule principale et l'atome lourd. Dans le premier cas, la zone d'intersection comportait un grand nombre d'invariants permettant de passer d'une ile à l'autre. Au contraire, maintenant, peu d'invariants assurent ce passage : quand un des ensembles moléculaires de l'unité asymétrique correspond à un atome lourd isolé, on observe dans l'espace réciproque des iles pratiquement disjointes.

III-1-3 Exemple d'application pratique

Un cas particulièrement inextricable de structures à atome lourd est celui où cet atome occupe une position spéciale dans la maille (miroir ou axe de symétrie). Bien souvent, dans ce cas, la structure de l'atome lourd est centrosymétrique, alors que l'ensemble de la structure ne l'est pas. Les phases des facteurs de structure calculées à partir de l'atome lourd sont toutes $0, \pm \frac{\pi}{2}, \pi$ et il est impossible en les affinant par la formule de la tangente d'obtenir une détermination différente.

La structure du bromhydrate de Dimidium (C.Courseille) (18) correspond pratiquement à ce cas et le processus décrit au début de ce paragraphe a permis de la résoudre sans grandes difficultés.

Elle appartient au groupe orthorhombique $Pna2_1$ (Z = 4) et la molécule principale a pour formule chimique :

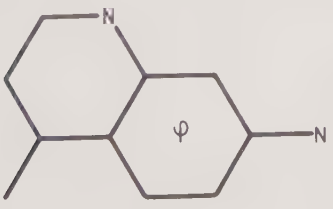


L'atome de brome fut localisé sur la fonction de Patterson en position presque particulière : x = 0,25 ; y = 0,07 ; z = 0,00 (arbitraire). De ce fait, les phases obtenues, à ce niveau, sont toutes voisines d'un "pôle" et la transformée de Fourier correspondante ne laisse apparaître aucun pic significatif.

On a alors entrepris l'étude de cette structure par les méthodes directes sans se soucier des résultats apportés par le Brome. Le processus de convergence choisit pour origine les réflexions 11 5 0 , 5 10 0 et 2 11 1. On fut obligé d'utiliser 2 symboles : 202 (a) et 405 (b). La réflexion 11 5 0 étant peu influencée par le Brome, nous l'avons assignée provisoirement m (0 ou π) pour ne pas, par un choix arbitraire, définir une origine différente de celle prise sur la fonction de Patterson.

La génération symbolique de phases montra que 5 10 0, 2 11 1 et 202 (a) d'une part, et 11 5 0 (m) et 405 (b) d'autre part définissaient deux ensembles de réflexions fortement liées. Par contre il n'existe pratiquement aucune relation entre ces deux ensembles : ils sont complètement disjoints. En comparant ces phases avec celles obtenues pour l'atome de Brome, nous avons pu identifier l'ensemble de corrélation (a) à l'ile du Brome et par conséquent l'ensemble de corrélation (b) représente l'ile du Dimidium.

Nous avons donné au symbole b différentes valeurs comprises entre 0 à 360° et nous avons choisi la solution qui, pour la zone d'intersection ($Fc_{Br} \sim \frac{1}{2} Fo$) donnait des valeurs de phases s'accordant le mieux avec celles fournies par le Brome.



Les sections de densités faites avec cette solution permirent de placer 12 des atomes de la molécule de Dimidium :
(Le reste des atomes furent placés par sections de Fourier successives).

III-2 ETUDE DU CAS GENERAL

(Cas où l'on ne connait absolument rien de la structure)

Ne connaissant aucun élément de la structure on ne peut savoir quelles réflexions appartiennent à un ensemble de corrélation ou à l'autre; nous devons donc essayer de définir les iles à priori. Ce type de résolution nous oblige à modifier les techniques de recherche employées jusqu'ici.

III-2-1 Choix des symboles

La première difficulté rencontrée est de choisir au moins un symbole dans chacune des iles moléculaires de façon à pouvoir par la suite contrôler le passage d'une molécule à l'autre et d'éviter ainsi les phénomènes de translation décrite au chapitre précédent (II-1).

Pour cela, on utilise le processus de convergence normal mais en ne considérant que les relations $\sum_2$ pour lesquelles la valeur A est élevée (A > 3 ou 4) évitant ainsi la zone d'intersection. Comme la sphère de basse résolution ne fournit que peu de relations invariantes, on est obligé pour générer par ce processus l'ensemble des phases de l'espace, d'avoir choisi au moins un symbole dans chacune des iles moléculaires.

En réalité, comme c'était d'ailleurs le cas quand il n'y avait qu'une seule molécule, on définit de cette façon plus de symboles qu'il n'est physiquement nécessaire. La première étape du calcul consistera à ramener cet ensemble de symbole aux seuls symboles "caractéristiques" des iles. Pour cela on peut utiliser l'addition symbolique mais il sera alors nécessaire d'examiner les relations entre symboles avec le plus grand soin. Il ne faut surtout pas vouloir que l'addition symbolique fournisse une seule solution : du fait de l'existence d'une zone d'intersection où les invariants de structure différent de zéro, on est assuré que la solution la plus probable fournie par les relations entre symboles ne correspond pas à la structure réelle.

Pour pouvoir réduire l'ensemble des symboles dans ce cas, nous avons introduit des figures de mérite analogues à celles utilisées précédemment. Mais comme elles interviennent avant l'affinement par la formule de la tangente, nous les appelerons "figures de mérite à priori".

III-2-2 Figures de Mérite à priori

Nous avons vu au Chapitre III de la 2ème partie que les différentes solutions obtenues dans la méthode de la multisolution pouvaient être classées en utilisant des figures de mérite (S_α ou Ψ_o) correspondant à des sommes du type :

$$\sum_H |\alpha_H|$$

où

$$\alpha_H^2 = \left| \sum_K A_{H,K} \cos \Phi'_K \right|^2 + \left| \sum_K A_{H,K} \sin \Phi'_K \right|^2$$

(nous désignons par Φ'_K la somme $\Phi_K + \Phi_{H-K}$)

En développant les deux carrés, on obtient, après avoir regroupé les termes carrés et les termes rectangles :

$$\alpha_H^2 = \sum_K A_{H,K}^2 + 2 \sum_K \sum_{K'} A_{H,K} A_{H,K'} (\cos \Phi'_K \cos \Phi'_{K'} + \sin \Phi'_K \sin \Phi'_{K'})$$

$$= \sum_K A_{H,K}^2 + 2 \sum_K \sum_{K'} A_{H,K} A_{H,K'} \cos \Delta\Phi_{KK'}$$

avec $\Delta\Phi_{KK'} = \Phi'_K - \Phi'_{K'}$

Le premier terme est essentiellement positif, α_H est maximum si le deuxième terme est maximum (tous les $\Delta\Phi_{K,K'}$ égaux à 0°).

Si on utilise comme figure de mérite $S_\alpha 2$ au lieu de S_α, nous aurons :

$$S_\alpha 2 = \sum_H \alpha_H^2 = \sum_H \sum_K A_{H,K}^2 + 2 \sum_H \sum_{K=K'} \sum A_{H,K} A_{H,K'} \cos \Delta\phi_{KK'}$$

Les différences $\Delta\Phi_{KK'}$, étant définies à partir de valeurs symboliques, elles ne pourront prendre qu'un nombre fini de valeurs possibles, si pour une valeur $\Delta\Phi_i$ particulière, on désigne par W_i le terme $(\sum_H \sum_{K=K'} \sum A_{H,K} A_{H,K'})$, la deuxième partie de $S_\alpha 2$ est égale à $2 \sum_i W_i \cos \Delta\Phi_i$.

Finalement $S_\alpha 2$ sera maximum ou minimum si $\sum_i W_i \cos \Delta\Phi_i$ est maximum ou minimum. (Notons que Schenk a récemment (1971) (32) proposé une fonction analogue en partant de figures de mérite différentes et en faisant ensuite un certain nombre d'approximations).

Cependant l'utilisation de $S_\alpha 2$ au lieu de S_α présente un inconvénient : cette nouvelle figure de mérite aura tendance à maximiser les réflexions ayant une valeur α élevée au dépens des autres. Afin d'éviter cet inconvénient on prendra comme figure de mérite la fonction $\sum_H \frac{\alpha_H^2}{\alpha_{H_{est}}}$, qui, tout en restant d'une utilisation aussi facile que $S_\alpha 2$ présente les mêmes avantages que S_α.

A partir des phases définies symboliquement on pourra utiliser deux figures de mérite analogues à celles utilisées dans la multisolution :

$$S_\alpha^* = \sum_H \frac{\alpha_H^2}{\alpha_{H_{est}}}$$

$$\Psi_o^* = \sum_{\substack{H \\ \text{non} \\ \text{observé}}} \frac{\alpha_H^2}{\alpha_{H\ \text{estimé}}}$$

c'est à dire maximiser ou minimiser des expressions de la forme $\sum_i W_i \cos\Delta\Phi_i$.

Il faut cependant reconnaître que le test Ψ_o^* ne donnera pratiquement aucun résultat : les réflexions non observées correspondent principalement aux zones de faible densité électronique c'est à dire à l'espace intermoléculaire. Pour avoir les renseignements sur les sommes α_H de ces réflexions il faut "connaître" les atomes situés à la périphérie des molécules. Or nous avons vu au début de cette deuxième partie que ceci n'était réalisé qu'en fin de génération de phases;et non pas au niveau de l'étape symbolique où l'on n'a déterminé qu'un petit nombre de phases pour des réflexions ayant une valeur E élevée.

III-2-3 Nouvelles considérations sur les figures de mérite (à posteriori)

Parmi les figures de mérite introduites dans la multisolution, 3 traduisent la cohérence de la détermination : S_α , F_{abs} et R_{Karle}: Pour une solution erronée elles prennent des valeurs abérrantes ($F_{abs} < 0,7$ et $R_{Karle} > 0,6$ par exemple). Cependant, dès l'instant où il apparait dans l'unité asymétrique plusieurs molécules différentes il se produit deux phénomènes contradictoires qui vont influencer les figures de mérite:

- d'une part à l'intérieur de chaque ile on doit avoir une détermination cohérente, c'est à dire que l'on ne rencontre pratiquement que des invariants nuls. A l'intérieur de chaque ile la somme S_α est maximale et le facteur résiduel de Karle minimum.

- d'autre part dans la zone d'intersection on est certain de rencontrer des invariants différents fortement de zéro. La valeur réelle de la somme S_α étendue à cette zone sera sûrement plus faible que la valeur maximale construite en considérant les invariants comme quasi-nuls. De même, le facteur résiduel de Karle pour ces réflexions pourra présenter certaines anomalies.

Finalement la somme S_α doit être suffisamment élevée pour que la cohérence à l'intérieur de chacune des iles soit parfaitement assurée, mais elle ne doit pas avoir une valeur trop grande qui indiquerait que dans la zone d'intersection on a identifié un grand nombre d'invariants à zéro, procédé qui entraine un "éclatement" d'une des 2 molécules, comme nous l'avons décrit au paragraphe II-1 de cette 2ème partie.

Le choix de la bonne solution dans un pareil cas est plutôt malaisé : dans l'ensemble des solutions classées par S_α croissant, la bonne solution ne correspond ni aux valeurs les plus fortes, ni aux valeurs les plus faibles mais à une valeur S_α moyenne.

Nous pensons qu'il est cependant possible de trouver cette solution, sans trop d'erreurs. Si on considère les solutions les mieux classées sur les figures de mérite "à priori" on est assuré d'obtenir une détermination cohérente à l'intérieur de chaque ile.

On peut alors remarquer, ensuite, que le calcul de la distribution théorique des cosinus invariants a été établi par H.Hauptman sans faire d'hypothèses sur la nature de la structure, donc indépendamment de l'existence d'iles moléculaires. La valeur $S_{\alpha est}$ qui dérive de cette distribution constitue donc une bonne estimation de la valeur réelle de S_α quelle que soit le type de structures étudiées (c'est à dire quelle que soit le groupe spatial et indépendamment de la présence de plusieurs éléments moléculaires dans l'unité asymétrique).

Dès lors la bonne combinaison de symboles correspond à une valeur S_α voisine de $S_{\alpha est}$. Généralement, du fait du rôle maximisant de la formule de la tangente, les solutions qui ont été définies à partir des figures de mérite à priori et qui correspondent à des iles moléculaires cohérentes, fournissent après affinement des valeurs S_α plus elevées que la valeur attendue $S_{\alpha est}$. Dans le cas de systèmes présentant le phénomène d'ile il est peut être plus judicieux de choisir, à ce niveau, une figure de mérite du type $S_\alpha - S_{\alpha est}$. Cette différence doit être minimale.

Le classement des différentes solutions peut donc être réalisé en utilisant comme test $S_\alpha - S_{\alpha est}$, R_{Karle} et surtout la fonction Ψ_o de Cochran.

III-2-4 Applications Pratiques

Nous avons rencontré, au cours de nos déterminations, deux structures présentant le phénomène d'iles moléculaires : l'oestrone III où il y a 2 molécules indépendantes dans l'unité asymétrique, et le complexe diéthylstilboestrol + ℓ-tyrosine éthyl ester. Ces deux structures appartenant au groupe spatial $P2_1$, il y a en plus un phénomène d'iles de groupe (que nous décrirons au chapitre suivant) mais il semble que l'influence des iles moléculaires soit prépondérante et ces deux structures peuvent parfaitement servir d'illustration au problème que nous venons d'étudier.

a) Détermination de la structure de l'oestrone III $P2_1$ $Z = 4$

Les deux molécules contenues dans l'unité asymétrique sont identiques, donc de même volume, et il existe de nombreux contacts entre elles. Par la suite nous n'observons pas d'iles absolument définies comme dans le cas de la structure du bromhydrate de dimidium. La zone d'intersection comprend déjà un grand nombre de réflexions et rares sont les réflexions qui ne sont influencées que par une seule molécule. De ce fait on ne peut être sûr que les symboles définis par le processus de convergence pourront

servir à caractériser les différentes iles. Pour essayer qu'il en soit ainsi, il vaut mieux prendre comme symboles, des réflexions ayant des indices élevés, c'est à dire situées hors de la sphère de basse résolution.

L'origine a été déterminée par le processus de convergence :

108	E =	2,14	0°
$\bar{2}11$	=	2,48	180°
$\bar{3}09$	=	4,02	0°

La phase de la réflexion 400 (E = 3,40) a été calculée à l'aide de la relation $\sum_1$.

Nous avons choisi trois réflexions comme symboles :

3 17 1	E = 3,03	a
0 5 6	= 2,13	b
1 17 2	= 3,59	c

Il apparut très vite la relation forte c = a+p où p est égal à $\pm \frac{\pi}{2}$. (Cette relation qui assure vraisemblablement la cohérence dans une des iles, devait s'avérer, une fois la structure terminée, exacte à 2° près).

Les relations obtenues par la suite étaient plus difficilement interprétables ; nous avons penser qu'elles correspondaient à la liaison entre les deux iles et qu'il valait sans doute mieux ne pas les résoudre de façon rigoureuse :

$$7b = p + 2a$$

$$4b = p + a$$

$$4b = \pi + a \quad \text{(relativement plus faible)}$$

(Notons que les 2 dernières relations donnent $p = \pi$ qui n'a pas de sens puisque nous savons que $p = \pm \pi/2$).

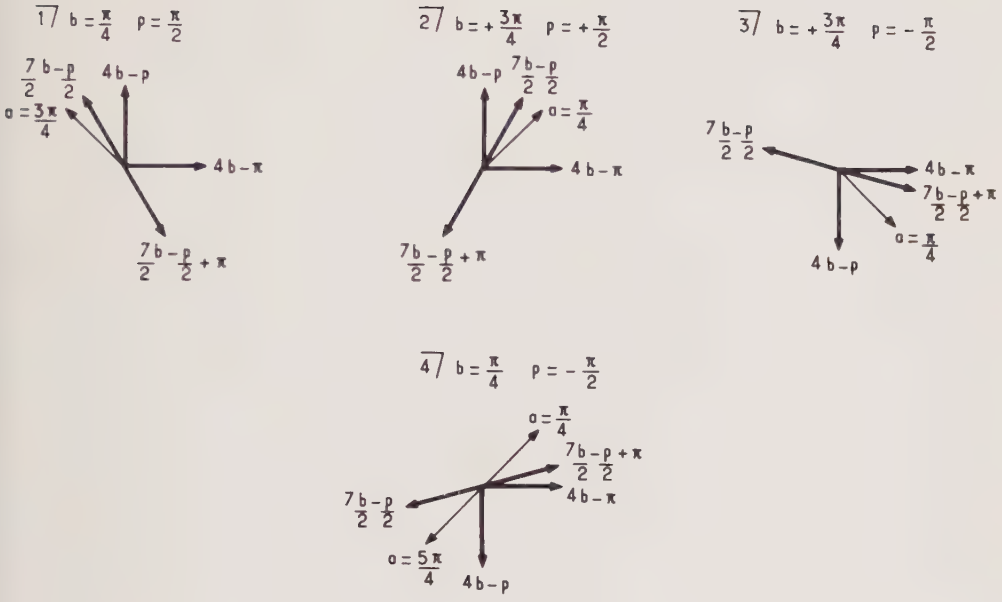
Pour éviter les inconvénients apportés par la définition "modulo 2π" de ces relations il vaut mieux les résoudre par rapport à a :

$$a = \frac{7}{2} b - \frac{p}{2} + k\pi$$

$$a = 4b - p$$

$$a = 4b - \pi$$

Nous avons utilisé b pour fixer l'énanthiomorphe, c'est à dire nous ne donnons à b que des phases positives $+ \pi/4$ et $+ 3\pi/4$. Les valeurs de a correspondantes se déduisent des schémas de la page suivante :



Suivant la valeur de b et p, il existe 8 valeurs de a possibles, nous en avons éliminées 3 pour lesquelles la variance angulaire était trop forte. (Nous n'avons pas utilisé pour a la valeur moyenne que l'on pourrait calculer à partir des schémas précédents, mais la valeur $(2k+1)\frac{\pi}{4}$ la plus proche).

Les 5 solutions restantes ont été affinées par la formule de la tangente; dans le tableau ci-dessous nous avons indiqué les valeurs correspondantes des figures de mérite utilisées (S_α, Ψ_o, R_{Karle}, $S_\alpha - S_{\alpha est}$), pour ces différentes solutions et le total n des places obtenues sur les tests suivants :

S_α, Ψ_o et R_{Karle} d'une part

$S_\alpha - S_{\alpha est}$, Ψ_o et R_{Karle}, d'autre part

		S_α	Ψ_o	R_{Karle}	$S_\alpha - S_{\alpha est}$	$n_{somme\ S_\alpha}$	$n_{somme\ S_\alpha - S_{\alpha est}}$
1	$b = \frac{\pi}{4}\ a = \frac{3\pi}{4}\ c = \frac{5\pi}{4}$	11.100	579	36,79	275	12	13
2	$b = \frac{\pi}{4}\ a = \frac{\pi}{4}\ c = -\frac{\pi}{4}$	11.032	503	36,63	157	8	4 x
3	$b = \frac{\pi}{4}\ a = \frac{5\pi}{4}\ c = \frac{3\pi}{4}$	11.043	514	36,68	168	9	7 x
4	$b = \frac{3\pi}{4}\ a = \frac{\pi}{4}\ c = \frac{3\pi}{4}$	11.232	572	36,10	357	6 x	10
5	$b = \frac{3\pi}{4}\ a = -\frac{\pi}{4}\ c = \frac{5\pi}{4}$	11.120	520	36,82	245	10	11

En utilisant les figures de mérite habituelles, la meilleure solution est la n°4. Des sections de Fourier obtenues à partir des phases correspondantes, laissent entrevoir malgré de nombreux pics parasites deux ensembles moléculaires ayant l'aspect des molécules d'oestrone. Le facteur résiduel sur la projection Oyz, après affinement, avec une agitation thermique isotrope, a pu être descendu jusqu'à 0,21 : la projection de la structure est correcte, mais cette solution n'est pas affinable dans l'espace. Si l'une des molécules était correcte, la deuxième était passablement déformée. Les valeurs attribuées aux symboles, qui donnent une valeur maximale pour S_α et sont les plus probables d'après les relations entre symboles (la solution 4 qui correspond au schéma n°2 précédent, correspond en effet à la variance angulaire la plus faible), ne fournissent pas la bonne solution. Elles tendent à rendre nuls les invariants de structure de la zone d'intersection, et désavantagent l'ile moléculaire "2" au profit de la première. Ce dernier point est encore plus évident si l'on compare ces valeurs, aux phases des facteurs de structure calculés à l'aide de la molécule "1" pour les 3 symboles :

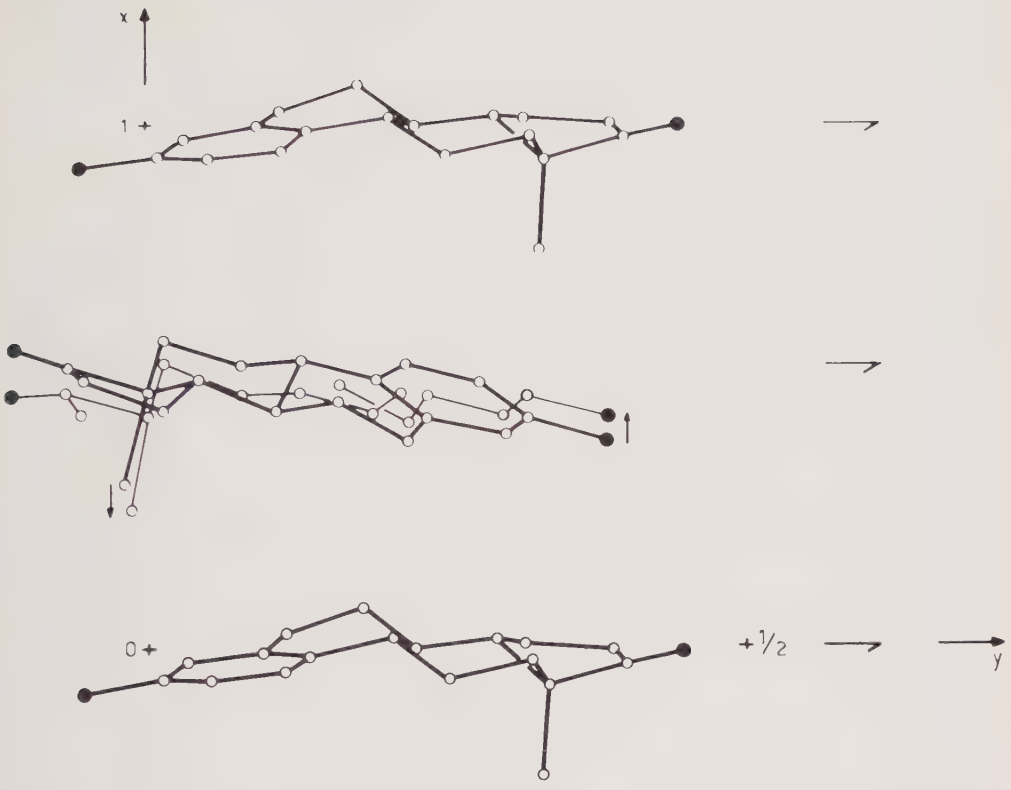
0 5 6 $\Phi = 105°$ $\Phi_{mol_1} = 137°$ $b = 135°$ $\Delta\Phi = 30°$

1 17 2 $\Phi = 165°$ $\Phi_{mol_1} = 169°$ $c = 135°$ $\Delta\Phi = 30°$

3 17 1 $\Phi = 73°$ $\Phi_{mol_1} = 89°$ $a = 45°$ $\Delta\Phi = 28°$

Remarque importante

Les sections de densités obtenues pour cette solution montre que l'éclatement de la "molécule 2" est dirigé par les contacts les plus courts existant entre les 2 molécules. L'affinement par la formule de la tangente tend à raccourcir ces contacts : le cycle A de la molécule 2 est "attirée" par la molécule 1 supérieure tandis que le méthyle attiré par le cycle A de la molécule inférieure entraine les cycles D et C :



Si nous utilisons au contraire la différence $S_{\alpha} - S_{\alpha est}$ au lieu de S_{α}, la solution 2 se détache nettement des autres. Cette solution fournit sans difficultés la structure cherchée, l'écart moyen entre les phases fournies par cette solution et celles obtenues en fin d'affinement ($R = 0{,}06_1$ (12a)) est d'ailleurs égal à $< \Delta \Phi > = 34°3$, ce qui est à peu près ce qu'obtient H.Hauptman pour la structure de l'oestriol qui présente les mêmes difficultés.

Les valeurs initiales des symboles étaient beaucoup plus proches de la réalité que précédemment : (la phase Φ réelle utilisée ici tient compte de la translation correspondante de la molécule parallèlement à Oy) :

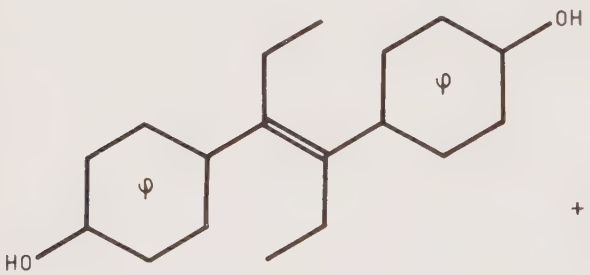
0 5 6	$\Phi = 30°$	b = 45	$\Delta\Phi = 15°$
1 17 2	$\Phi = 150°$	c = 135°	$\Delta\Phi = 15°$
3 17 1	$\Phi = 242°$	a = 225°	$\Delta\Phi = 17°$

La solution n°3 qui occupe alors la deuxième place dans le classement donné par les figures de mérite ne diffère pas réellement de la solution précédente ; on pouvait s'en douter en remarquant que les valeurs numériques prises par les figures de mérite sont voisines (les deux déterminations présentent la même cohérence) alors que les valeurs des phases obtenues sont totalement différentes. Ces deux solutions représentent effectivement la structure réelle mais correspondent à des origines différentes sur l'axe 2_1. On passe d'une structure à l'autre par translation parallèle à b. Ce phénomène est fréquent dans les groupes où le choix de l'origine est arbitraire dans une direction.

Bien que l'on croit avoir fixé l'origine au départ (ici à l'aide de la réflexion $\bar{2}11$), du fait que les symboles définis par la suite n'ont pas une valeur exacte, on assiste au cours de l'affinement par la formule de la tangente à une migration de l'origine le long de l'axe 2_1, pour que l'ensemble des phases constitue le groupe le plus homogène possible. Il vaut donc mieux permettre à la réflexion utilisée pour fixer l'origine dans cette direction de varier au cours de l'affinement: la valeur finale de sa phase donnera une idée de la migration d'origine qui a pu se produire.

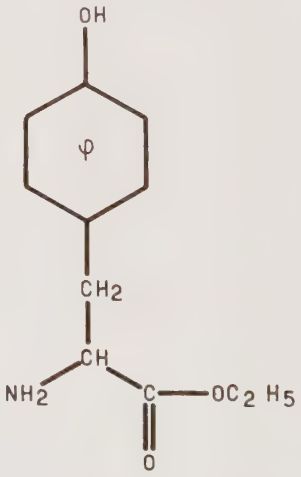
b) Détermination de la structure du complexe DES + l-tyrosine éthyl éther $P2_1$ Z=2 (10)

Si ici, nous avons encore dans l'unité asymétrique deux molécules de même importance, elles sont de natures chimiques différentes :



DES

\+



L'origine définie par le processus de convergence est :

5 0 $\bar{6}$	E = 2,30	0°
6 1 6	E = 2,40	0°
5 0 7	E = 3,99	0°

Du fait du phénomène d'iles moléculaires (et d'iles de groupe) nous avons ,ici encore, besoin d'un nombre de symboles plus élevé que d'habitude :

$\bar{5}$ 3 3	2,75	a
$\bar{4}$ 6 5	2,18	b
9 3 3	1,97	c
$\bar{2}$ 7 4	2,17	d

Si on voulait utiliser la multisolution dès le départ on serait obligé d'étudier $4^4/2 = 128$ solutions différentes (il faut tenir compte du choix de l'énanthiomorphe).

Nous avons utilisé dans une première étape l'addition symbolique pour tenter de réduire le nombre de solutions. Les valeurs de symboles définies à $(2n+1)\frac{\pi}{4}$ ou sur un "pole" ont été testées à l'aide de la figure de mérite décrite au paragraphe III.2.2 :

$$\sum_i W_i \cos \Delta\Phi_i$$

Le terme $\sum_i W_i$ est égal à 356,03. La valeur de la figure de mérite pour différentes valeurs des symboles est donnée par :

$$\sum_i W_i \cos\Delta\Phi_i = 127{,}84\cos|2c+4(b-d)| + 83{,}10\cos|2c+6(b-d)| + 60{,}95\cos 2(a+3b+2c-4d)$$

$$+ 16{,}32 \cos 2(b-d) + \ldots$$

Nous avons omis ici 16 termes ayant un poids inférieur à 8,0. Dans cette détermination, les 4 termes de $S_\alpha 2$ les plus importants représentent 81% de la valeur totale. On peut choisir les solutions qui rendent ces 4 termes maximaux pour utiliser la multisolution : il ne reste plus alors que 16 cas à étudier. Comme nous l'avions prévu, le test en Ψ_0 appliqué à ce stade ne donne aucun renseignement utilisable, la valeur $\sum_i W_i$ maximale étant beaucoup trop faible (de l'ordre de 5,0 alors qu'en fin d'affinement le test en Ψ_0 donne des sommes sur α de l'ordre de 500,0).

Le facteur résiduel de Karle ne donnant pas des résultats suffisamment contrastés nous n'avons utilisé en fin d'affinement que les tests S_α et Ψ_0 pour classer les solutions.

la solution $a = 3\pi/4$, $b = 3\pi/4$, $c = 0$, $d = 3\pi/4$ apparait alors de loin la meilleure : 1er sur Ψ_0 (341,8) et 3ème sur S_α (0,9257)

la solution suivante est :

2ème sur Ψ_0 (480,7) et 5ème sur S_α (0,8669)

Les écarts importants observés sur les figures de mérite nous ont fait penser que cette première solution était sûrement la bonne. Effectivement, des sections de densités obtenues à partir de cette solution, permettent de placer les 36 atomes semi lourds de la structure sans difficultés.

III-3 CONCLUSION

Quiconque a déjà déterminé des structures cristallines, sait parfaitement que certaines réflexions de l'espace réciproque ne sont influencées que par un groupe d'atomes voire même par un atome unique. Nous avons montré que l'ensemble des réflexions influencées par une molécule s'identifiait à un ensemble de corrélation, c'est à dire que toutes ces réflexions étaient reliées entre elles par des relations $\sum_2$ fortes.

Nous avons montré également que dès l'instant où il existait différentes molécules dans l'unité asymétrique, les relations entre symboles ne pouvaient plus être utilisées avec la même certitude qu'auparavant et qu'assurèment la solution la plus probable c'est à dire celle qui correspond à un S maximum ne pouvait pas être la solution correcte. Ces constatations nous ont amenées à changer nos méthodes d'investigations et surtout à utiliser des figures de mérite différentes. Nous pensons également qu'il devrait être possible après cette étude, de trouver une structure correcte à partir de molécules éclatées en essayant de déterminer quelles translations Δr_o ont été causes de cette déformation.

Avant de terminer cette étude des méthodes directes nous voudrions analyser le deuxième phénomène pouvant entrainer l'existence d'ensembles de corrélation dans l'espace réciproque : la nature du groupe spatial.

C H A P I T R E IV

ILES DE GROUPE

Au cours des chapitres précédents, nous avons vu que l'existence d'ensembles moléculaires distincts dans l'unité asymétrique créait, au niveau de l'espace réciproque, des sous ensembles de corrélation ou iles.

En réalité, l'apparition d'iles peut avoir une cause différente : la nature du groupe de symétrie. On rencontre ainsi des sous ensembles de corrélations pour les groupes centrosymétriques ne présentant pas d'éléments de translation ($P\bar{1}$, P2/m). Pour les groupes non centrés (P1, P2), le phénomène observé est légèrement différent. Malgré la présence d'un axe hélicoïdal, le groupe spatial $P2_1$ présente également des particularités assez semblables.

Nous voudrions montrer dans ce chapitre, que les iles qui nous intéressent maintenant, sont de nature totalement différente de celles que nous venons d'étudier.

IV-1 LE PHENOMENE D'ILES DANS LE GROUPE $P\bar{1}$

IV-1-1 Existence d'iles de groupe

Si nous avons affaire à un tel groupe, nous avons obligatoirement au moins deux iles dans l'espace réciproque :

Les réflexions h_1, k_1, l_1 et $\bar{h}_1$, $\bar{k}_1$, $\bar{l}_1$ sont liées par symétrie, leurs facteurs de structures sont égaux :

$$E_{h_1 k_1 l_1} = E_{\bar{h}_1 \bar{k}_1 \bar{l}_1}$$

Considérons une des relations $\sum_2$ qui définit une réflexion H_2 à partir de H_1. Nous aurons :

$$s(E_{h_2 k_2 l_2}) = s(E_{h_1 k_1 l_1} \times E_{h_2-h_1, k_2-k_1, l_2-l_1})$$

(s( ) = signe de)

Appliquons la symétrie centrale au 2ème membre de l'équation :

$$s(E_{h_2 k_2 l_2}) = s(E_{\bar{h}_1 \bar{k}_1 \bar{l}_1} \times E_{h_2+h_1, k_2+k_1, l_2+l_1})$$

d'où nous pouvons déduire :

$$s(E_{h_2-h_1, k_2-k_1, l_2-l_1}) = s(E_{h_2+h_1, k_2+k_1, l_2+l_1})$$

Soit en posant $H_1 - H_2 = H$ $\quad s(E_{hkl}) = s(E_{h+2h_2, k+2k_2, l+2l_2})$ $\forall$ hkl

Cette relation montre que toutes les réflexions d'une même classe de parité ont même signe, dès l'instant où elles sont liées entre elles par des relations $\sum_2$ fortes. L'origine étant définie dans 3 classes de parité linéairement indépendantes, si on affecte un signe + à chacune des 3 réflexions origine, la relation ci-dessus implique que le signe de toutes les réflexions de l'espace est +. Comme cette dernière affirmation ne peut être exacte, il existe obligatoirement au moins 2 ensembles de corrélation disjoints, c'est à dire entre lesquels la relation $\sum_2$ n'est pas vérifiée : un ensemble (+) et un ensemble (-). Nous rencontrons déjà ici une première différence avec les iles moléculaires, où chaque ensemble renferme à la fois des réflexions + et - .

IV-1-2 Dépendance du choix de l'origine

Précédemment nous avons pu définir une ile "molécule 1" et une ile "molécule 2" et attribuer une réflexion donnée à l'une des iles. Ici, nous avons bien défini une ile (+) et une ile (-), mais pouvons-nous, pour autant, dire qu'une réflexion appartient à priori à l'une de ces iles ?

Le choix des signes des réflexions fixant l'origine est complétement arbitraire. Une fois déterminées les 3 réfléxions les plus appropriées

pour fixer l'origine, nous pouvons leur affecter à toutes 3 le signe + ; elles appartiennent alors à l'ile (+). Mais nous aurions très bien pu en fixer arbitrairement une négative et la faire entrer ainsi dans l'ile (-).

La constitution des iles de groupes dans le cas du groupe $P\bar{1}$ dépend donc du choix de l'origine. Alors que les iles moléculaires peuvent être considérées comme des invariants de structure et de ce fait être définies à priori, les iles de groupe ne peuvent être constituées qu'après avoir fait choix d'une origine.

IV-1-3 Schéma général des iles de groupe

Considérons maintenant les réflexions de l'ile (+). Pour simplifier le problème nous supposerons que la structure ne comporte qu'un seul atome de coordonnées x, y, z.

Une réflexion H(hkl) de cet ensemble de corrélation vérifie dans le groupe $P\bar{1}$, la relation $hx+ky+lz = n$ où n est en entier.

soit H'(h'k'l') une autre réflexion de cet ensemble : $h'x+k'y+l'z = n'$

Alors la réflexion H-H' reliée par une relation $\sum_2$ aux réflexions H et H' appartient également à l'ile (+) car nous avons :

$(h-h')x + (k-k')y + (l-l')z = n-n'$ qui est aussi un entier.

L'invariant de structure H, H' et H-H' est positif. Les diverses réflexions de l'ile (+) sont donc liées entre elles par des relations $\sum_2$ fortes qui correspondent à des invariants positifs.

Considérons maintenant les réflexions de l'ile (-) et soit H et H' deux réflexions de cette ile :

$$hx + ky + lz = \frac{1}{2}(2n+1)$$

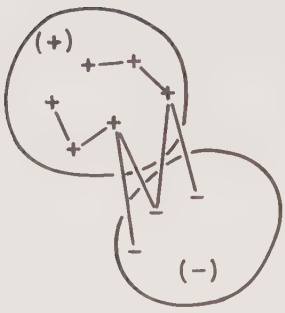
$$h'x+k'y+l'z = \frac{1}{2}(2n'+1)$$

Par relation $\sum_2$, elles génèrent la réflexion H-H' pour laquelle nous aurons :

$(h-h')x + (k-k')y + (l-l')z = n-n'$ qui est un entier.

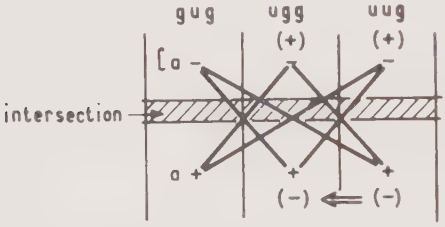
Cette réflexion n'appartient donc pas à l'ile (-), mais à l'ile (+). Par la suite les différentes réflexions de l'ile (-) ne sont pas liées entre elles par des relations $\sum_2$. Elles ne peuvent être générées que par l'intermédiaire des réflexions de l'ile (+).

L'ile (-) que nous venons de décrire ne correspond donc pas à la définition que nous avions posée au départ : celles d'un ensemble de réflexions reliées les unes aux autres par des invariants de structure positifs.



Il résulte de l'étude ci-dessus que s'il n'existait qu'un seul atome dans l'unité asymétrique l'origine étant définie arbitrairement à l'intérieur de l'ile (+), il n'existerait absolument aucune possibilité de passer dans l'ile (-). Dans ce cas très particulier, les ensembles de corrélation sont parfaitement disjoints (même en considérant toute relation $\sum_2$, aussi faible soit elle). On pourrait, à l'aide des relations $\sum_2$ explorer complètement l'ile (+), puis ensuite attribuer le signe - aux réfléxions restantes.

Dès l'instant où la structure comporte plusieurs atomes, il existera des possibilités de passage d'une ile à l'autre, car il existe des invariants de structure négatifs : les 2 iles ne sont plus complètement distinctes et présentent une intersection.



En fait, le schéma des iles de groupe donné ci-dessus ne correspond pas tout à fait à la réalité puisque nous n'avons pas fait intervenir la dépendance du choix de l'origine. Notamment, il n'existe pas 2 types d'invariants de structure (ceux qui joignent les réflexions + et ceux qui relient les réflexions - à travers l'ile (+) comme nous l'avons représenté ci-dessus. Les iles sont définies tout d'abord au niveau de chaque classe de parité où il existe 2 ensembles complémentaires a et Ca. Suivant le choix de l'origine (a = + ou a = -) on aura une définition différente des iles, et les différents invariants sont ainsi amenés à jouer le même rôle.

IV-1-4 Intersection

Alors qu'à propos du phénomène d'iles moléculaires, on pouvait parler de réflexions influencées à la fois par les 2 motifs atomiques et les attribuer tout naturellement à la zone d'intersection, dans le cas d'iles de groupes on ne peut imaginer des réflexions qui appartiendraient à la fois à l'ile (+) et à l'ile (-). Par suite, le passage d'une ile à l'autre ne peut s'effectuer par l'intermédiaire d'un ensemble-intersection constitué de réflexions.

La zone d'intersection est constitué par l'ensemble des invariants négatifs (dans le cas où il n'y a qu'un seul ensemble moléculaire). Si on a affaire, en plus, à des iles moléculaires il faudrait séparer les invariants négatifs correspondant aux deux phénomènes.

IV-2 PROBLEMES POSES PAR LES ILES DE GROUPE

Ce serait une erreur de croire que les difficultés rencontrées dans le cas de structure présentant des iles de groupe sont dues à une répartition non statistique des atomes dans la maille et notamment à une proportion plus importante d'invariants négatifs.

Il n'y a à priori aucune raison pour que l'on ait une répartition différente des cosinus invariants dans le groupe P2/n et le groupe $P2_1/c$. C'est uniquement la façon dont on utilise les invariants négatifs qui est la cause de l'échec des méthodes directes dans le cas de certains groupes de symétrie.

- Dans le groupe $P2_1/c$ on a un seul ensemble de corrélation. En effet, quand on change k en -k il y aura changement de signe si k+l = 2n+l c'est à dire pour les 4 classes de parité (gug, uug, ggu et ugu). Or il est obligatoire que l'une au moins des réflexions origine appartiennent à l'une de ces classes. Par suite, dès le départ, du fait de la symétrie, on aura défini une réflexion + et une réflexion - . Pour générer l'ensemble des réflexions de l'espace on n'est donc pas obligé de passer par les invariants négatifs de la zone d'intersection. Les relations entre symboles que l'on obtient alors pourront généralement être considérées comme exactes.

- Dans les groupes $P\bar{1}$, P2/n on a au contraire affaire à 2 iles distinctes et l'on ne pourra les définir qu'en faisant appel à la zone d'intersection. On retrouve dès lors les mêmes difficultés que nous avons rencontrées pour les iles moléculaires. Notamment on ne peut plus tenir compte des relations $\sum_2$ car elles rendent la somme S_α maximum et l'on sait que la bonne solution ne correspond pas à une valeur $^\alpha S$ maximum mais plutôt à la valeur que prévoit la théorie c'est à dire une valeur S_α proche du $S_{\alpha est}$.

Une mauvaise solution correspond à l'identification à zéro d'un certain nombre de phases invariantes non nulles, et revient à relier par liaison covalente des atomes qui ne le sont pas normalement. Par contre d'autres parties de la molécule sont éclatées. Même si on a la chance d'obtenir sur des sections de densités une molécule entière, celle-ci sera généralement translatée de façon à former des "dimères" au niveau d'un contact intermoléculaire prononcé.

La résolution de structures de ce type nécessite les mêmes précautions que celles décrites aux chapitres précédents : il vaut mieux utiliser au départ les invariants de structures les plus forts, et si on est obligé au cours de la génération d'utiliser une phase ayant une mauvaise détermination, il est préférable de la remplacer par un "symbole de contrôle" pour la déterminer avec le maximum d'informations. En effet, alors que pour un groupe classique ($P2_1/c$, Pbca, $P2_12_12_1$) une détermination erronée pourra être corrigée automatiquement au cours de la génération, il n'existe aucune possibilité de rattraper une détermination erronée quand on est en présence d'iles de groupe.

Remarque : Il est également vraisemblable que l'utilisation de la formule modifiée de la tangente qui n'identifie pas implicitement la phase invariante à zéro, s'avère préférable dans le cas d'iles de groupe, puisqu'elle ne devrait pas maximiser la zone d'intersection.

IV-3 EXEMPLES DE RESOLUTION

IV-3-1 Androstanolone, H_2O $P2_1$ Z = 2

Le processus de convergence permet de choisir pour origine, les réflexions :

$\bar{3}06$	3,12	0
$\bar{5}07$	2,61	0
110	1,64	0

La réflexion 200 fut déterminée π à l'aide de la relation $\sum_1$. Deux symboles furent utilisés :

232	2,96	b
237	2,54	a

En raison des considérations précédentes sur les iles de groupe, nous n'avons considéré comme significatives, que les deux relations entre symboles les plus fortes $4b = 0$ $(b = \pm \pi/2)$

$$\text{et } a = b + k\pi$$

b = 0 ou π ne peut être solution car dans ce cas toutes les phases seraient 0 ou π. Pour fixer l'énanthiomorphe nous avons considéré seulement la détermination positive de la phase b, et dans l'esprit de la multisolution nous avons finalement donné à cette phase 2 valeurs successives $b = + \pi/3$ et $b = + 2\pi/3$. 4 solutions différentes ont donc été étudiées.

Si l'on considère alors les figures de mérite obtenues, on s'aperçoit que S_α et Φ_o donnent des résultats contradictoires comme c'est souvent le cas dans des structures de ce type : ils sont respectivement maxima et minima en même temps. Généralement ce phénomène traduit un certain nombre d'imperfections dans la détermination; aucune des 4 solutions étudiées ne correspond à la structure réelle. Cependant elles peuvent néanmoins permettre de résoudre le problème :

- nous avons d'abord étudié la solution correspondant à la valeur S_α maximale ; elle fournit une molécule absolument correcte au point de vue conformation, cependant les contacts intermoléculaires sont aberrants (certains contacts de Van der Waals ont été réduits à la longueur d'une liaison covalente). Nous avons supposé que la conformation et l'orientation de la molécule obtenue était correcte et nous avons essayé de déterminer la translation nécessaire par minimisation du facteur résiduel dans le plan h0l. Ce procédé nous a alors facilement fourni la solution correcte ; elle n'est pas très éloignée de celle obtenue par les méthodes directes :

$$x = + 0,06$$

$$z = + 0,12$$

ce qui correspond à peu près à une translation de 1,5 A c'est à dire à la différence qui existe entre une liaison de covalence et un contact de Van der Waals. Notons par ailleurs que cette translation est parallèle à la direction générale de la molécule (qui correspond au plan $\bar{2}01$).

- la solution classée troisième pour le S_α (ou deuxième pour le Ψ_α) fournit une projection $h0\ell$ absolument correcte: les paramètres de translation y sont convenablement déterminés, par contre la molécule est éclatée dans la direction Oy.

IV-3-2 Testostérone, H_2O $P2_1$ Z = 2

L'exemple précédent montre qu'il est possible de résoudre une structure présentant un phénomène d'iles de groupe malgré une mauvaise détermination. La testostérone hydratée, qui est une structure très voisine, permet de voir que, si par un moyen quelconque on parvient à remettre la détermination de phases dans le droit chemin, tous les écueils rencontrés précédemment disparaissent.

L'origine fixée par le processus de convergence est :

409	4,32	0°
$\bar{3}01$	2,70	0°
718	2,28	0°

L'utilisation de 3 symboles est nécessaire pour amorcer la détermination :

608	3,05	m(0 ou π)
237	2,91	a
$\bar{3}16$	2,46	c

L'ordre d'apparition des réflexions est alors

	E	Φ	seuil d'acceptation ($\frac{1}{2}\alpha$)	
$\bar{1}$ 1 5	2,96	m+c	4,86	
7 1 3	2,52	π+c	7,35	
$\bar{1}$ 2 13	2,50	π+a-c	6,26	
1 0 10	2,63	0	4,55	
5 2 1	1,80	a	5,11	
2 0 10	2,48	m	4,82	
$\bar{5}$ 1 2	1,91	π+m	4,75	
6 1 8	2,28	m	1,70(*)	(donnée par la $\bar{5}$ 1 2 et la 1 0 10)
3 1 9	2,52	m	2,30	
$\bar{4}$ 3 9	1,94	a	2,01	
$\bar{4}$ 0 1	2,49	m	4,48	

On remarque, une brusque chute du seuil d'acceptation au niveau de la réflexion 618. Une telle détermination peut être erronée; comme elle va influer sur toutes les autres déterminations, nous avons préféré nous assurer de la phase de cette réflexion à l'aide d'un symbole de contrôle.

Nous avons ainsi mis en évidence que la phase correcte de la réflexion 618 était plutôt π+m, et avons réussi à obtenir une bonne détermination de phases; à ce stade une des solutions de la multisolution donne une valeur maximale sur S_α et minimale sur Ψ_0. Des sections de densité faites à l'aide de cette solution ont permis de placer tous les atomes de la molécule de testostérone. Cette molécule est de plus située correctement dans la maille.

CONCLUSION

Nous pensons par l'étude précédente avoir quelque peu modifié les idées classiques émises sur les méthodes directes de résolution de structure cristalline

Ces méthodes reposent au départ sur une distribution statistique des atomes dans la maille cristalline correspondant pour l'espace réciproque à un ensemble unique de réflexions fortement liées.

Les échecs des méthodes directes sont dus à l'existence d'exceptions à cette hypothèse générale liées notamment :

- à la nature du groupe
- à la présence de différents éléments moléculaires.

D'après nos résultats expérimentaux, il semblerait que le phénomène d'iles moléculaires soit nettement moins gênant pour la détermination de structures que le phénomène d'ile de groupe, qui d'ailleurs a tendance à s'atténuer à mesure que le nombre d'atomes dans la maille cristalline augmente :

Il nous est en effet finalement apparu plus facile de réaliser une structure de stéroïde dans le cas $P2_1$ Z = 4 que dans le cas $P2_1$ Z = 2.

On peut donc penser que les difficultés rencontrées dans le cas de petites structures doivent s'estomper à mesure que le nombre d'atomes dans la maille cristalline augmente. Une pareille affirmation est en nette contradiction avec l'idée couramment admise que les méthodes directes sont inutilisables pour la détermination des structures de très grosses molécules. On doit donc admettre qu'il existe à mesure que le nombre d'atomes augmente des difficultés de nature différente de celles que nous avons étudiées et l'exploration de cette limite supérieure de l'utilisation des méthodes directes nous semble indispensable...

A N N E X E

OBSERVATIONS EXPERIMENTALES

(Description sommaire des différentes structures cristallines déterminées :

- distances interatomiques, angles de valence, angles dièdres
- projections et distances intermoléculaires)

OESTRADIOL + Alc.PROPYLIQUE

B.BUSETTA, C.COURSEILLE, S.GEOFFRE et M.HOSPITAL - Acta Cryst.(1972) B 48 p.1349

$P2_12_12_1$ Z = 4 a = 12,215
b = 24,251 ± 0,003 V = 1976,12 $\AA^3$
c = 6,671

R final = 0,061 Détermination par minimisation directe de R (x,y,z,Θ, Ψ, Φ)

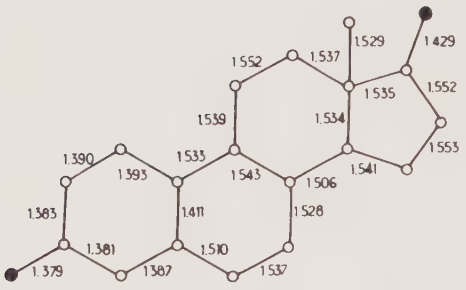
OESTRADIOL HEMIHYDRATE

B.BUSETTA et M.HOSPITAL
C.R.Acad.Sc.Paris t 268 p.1300 (1969)
Acta Cryst. (1972) B 28 p.560

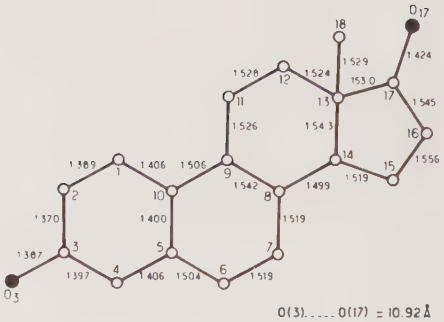
$P2_12_12$ Z = 4 a = 12,055
b = 19,280 ± 0,003 Å V = 1541,40 $\AA^3$
c = 6,632 d_{mes} = 1,20 $d_{calc.}$ = 1,21

R final = 0,065 Détermination par minimisation directe de R(x,y,z, Θ, Ψ, Φ)

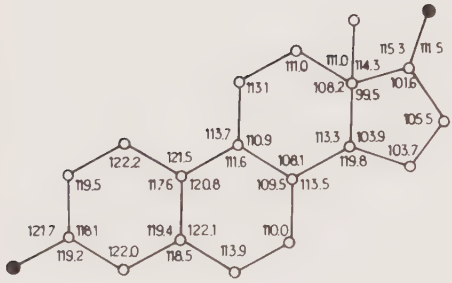
OESTRADIOL + Alc Prop.

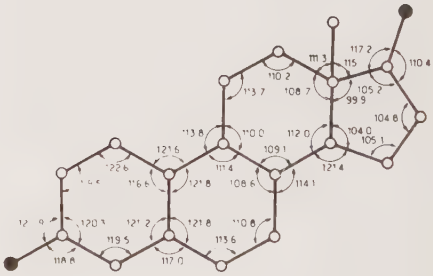


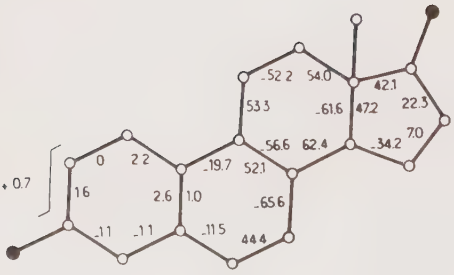
OESTRADIOL + H_2O

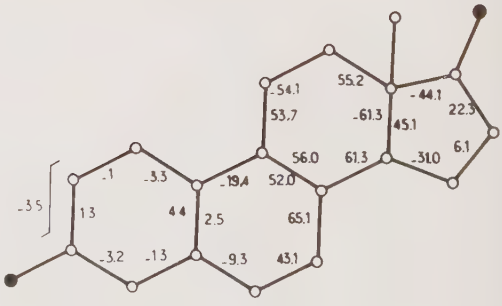


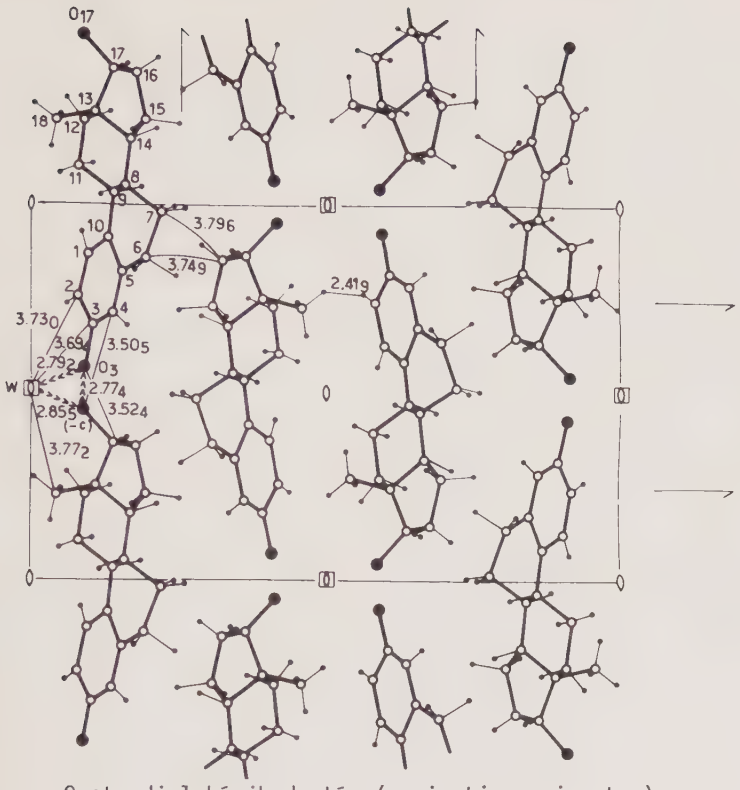
O(3)..... O(17) = 10.92 Å



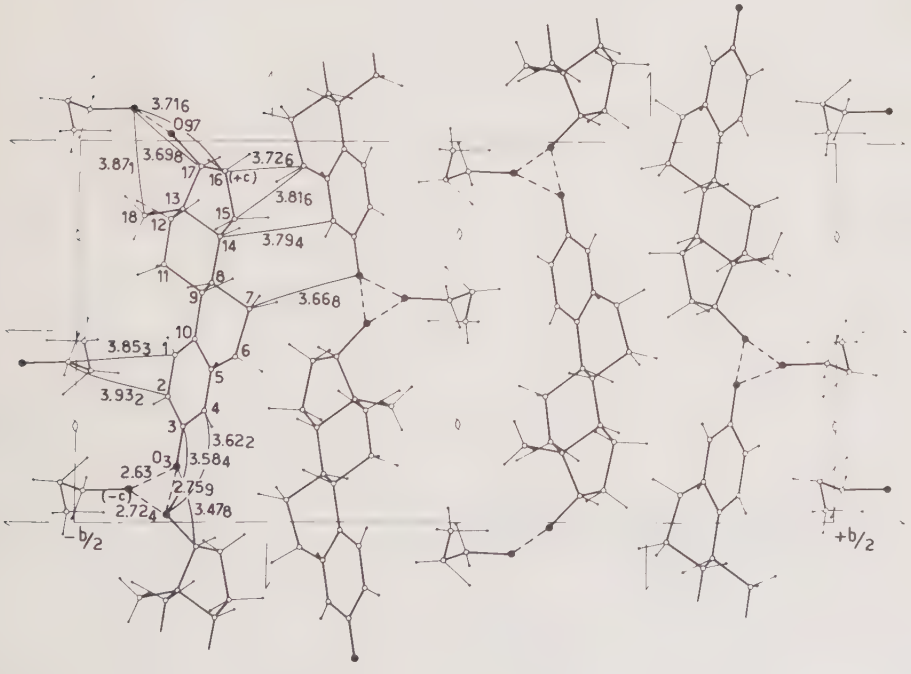








Oestradiol hémihydraté (projection suivant c)



Oestradiol + alc.propylique (projection suivant c)

OESTRONE SUBLIME (Forme I)

B.BUSETTA, C.COURSEILLE, M.HOSPITAL
C.R.Acad.Sc.Paris t.272 p.1211 (1971)

$P2_12_12_1$ Z = 4 a = 12,188
b = 16,301 ± 0,003 V = 1481,81 Å^3
c = 7,463

R_{final} = 0,069 Détermination par utilisation simultanée de l'addition symbo
de phases et de la multisolution.

OESTRONE (Forme II) cristallisé dans acétone

B.BUSETTA, C.COURSEILLE, M.HOSPITAL
C.R.Acad.Sc.Paris t.271 p.1558 (1970)

$P2_12_12_1$ Z = 4 a = 10,043
b = 18,424 ± 0,003 Å V = 1440,75 R^3
c = 7,787

R_{final} = 0,039 Détermination par méthode d'addition symbolique de phases.

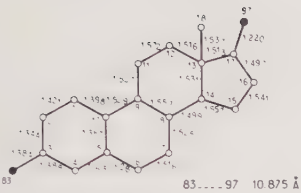
OESTRONE SUBLIME (Forme III)

B.BUSETTA, C.COURSEILLE, Mme F.LEROY, M.HOSPITAL
C.R.Acad.Sc.Paris t.274, p.153-156 (1972)

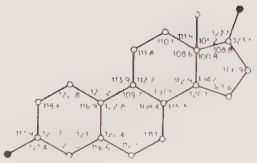
$P2_1$ Z = 4 a = $9,27_1$
b = $22,28_5$ ± 0,005 Å β = 111°27' ± 12' V = 1461
c = $7,61_0$

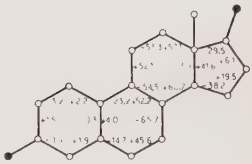
R_{final} = $0,06_1$ Détermination par la méthode de la multisolution et addition
symbolique.

OESTRONE I

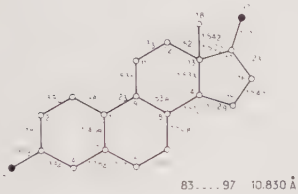


83....97 10.875 Å

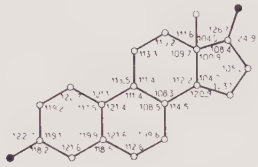


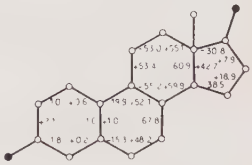


OESTRONE II

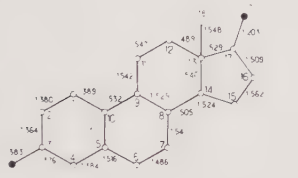


83....97 10.830 Å

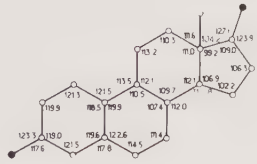


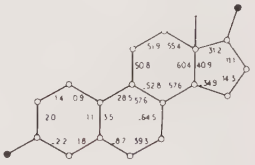


OESTRONE III A

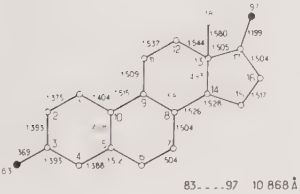


83....97 10.829 Å

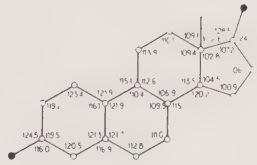


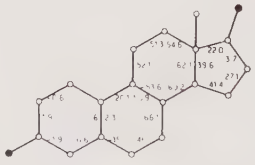


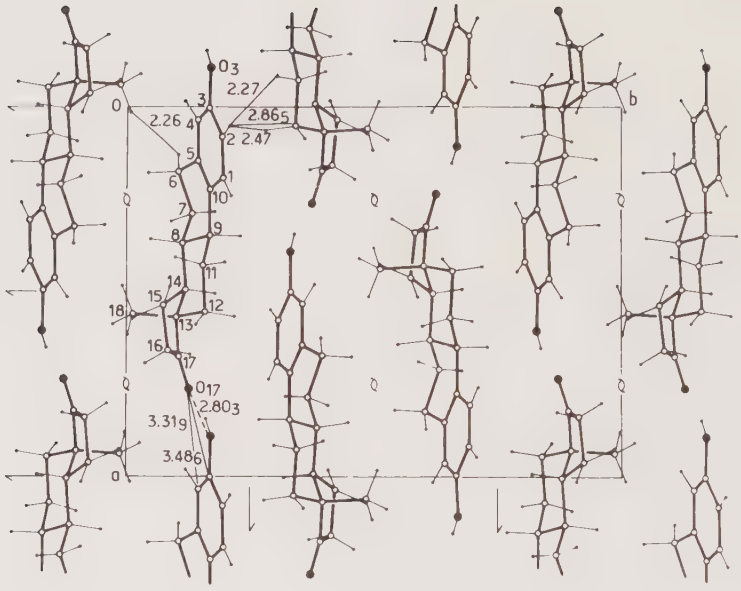
OESTRONE III B



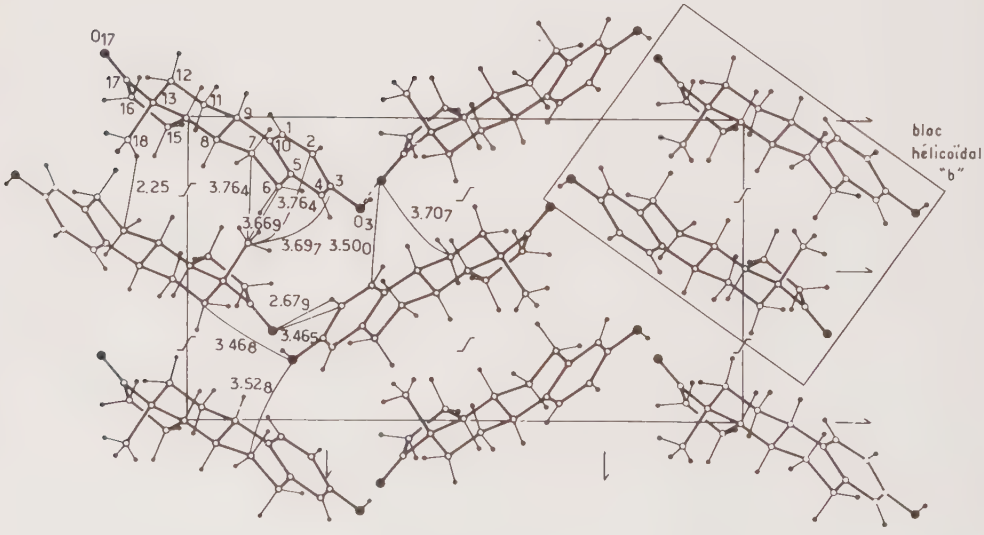
83....97 10.868 Å



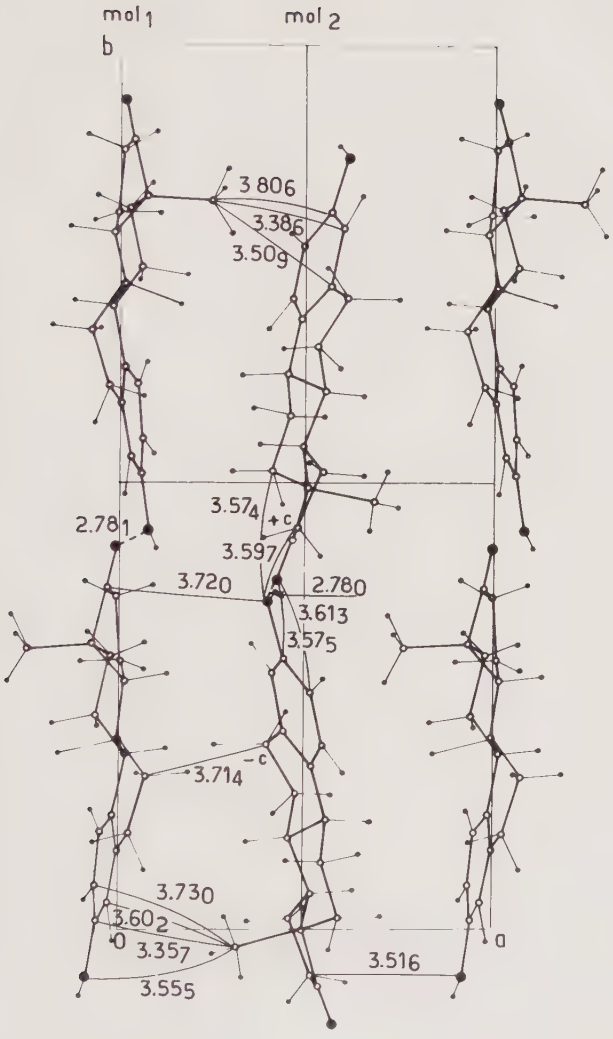




Oestrone I (Projection suivant c)



Oestrone II (Projection suivant c)



Oestrone III (Projection suivant c)

ANDROSTANOLONE + H_2O (5 α androstan 3 one 17β ol)

B.BUSETTA, C.COURSEILLE, J.M.FORNIES-MARQUINA et M.HOSPITAL
Cryst.Struct.Comm. (1972) 1 p.43.

$P2_1$ Z = 2 a = 9,406
b = 8,074 ± 0,003 Å β = 99°1
c = 11,549

$R_{final} = 0,05_3$ Détermination par utilisation successive des méthodes directes (add.symb.et multisolution) et de la minimisation sur les paramètres de translation R(x, y, z).

TESTOSTERONE + H_2O

B.BUSETTA, C.COURSEILLE et M.HOSPITAL
Acta Cryst. (1972) B 28 p.3293

$P2_1$ Z = 2 a = 9,602
b = 8,104 ± 0,003 Å β = 97°80 V = 859,3 $Å^3$
c = 11,155

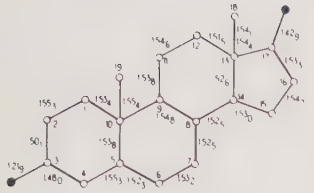
$R_{final} = 0,045$ Détermination par utilisation simultanée de l'addition symbolique et de la multisolution.

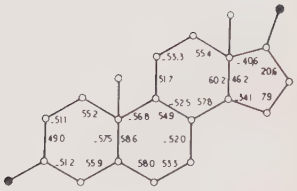
ANDROSTENEDIONE

B.BUSETTA, G.COMBERTON, C.COURSEILLE et M.HOSPITAL
Cryst.Struct.Comm. (1972) 1 p.129

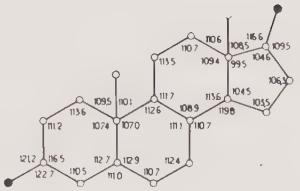
$P2_12_12_1$ Z = 4 a = 12,978
b = 16,949 ± 0,005 Å
c = 7,365

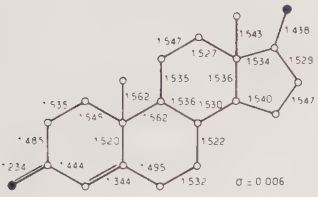
$R_{final} = 0,044$ Détermination par méthodes directes à l'aide des invariants de structure (évaluation par la formule du triple produit H.Hauptm
Acta Cryst. 1969 B 25 p.814).

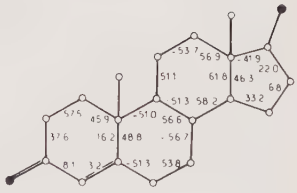




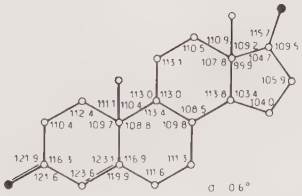
ANDROSTANOLONE

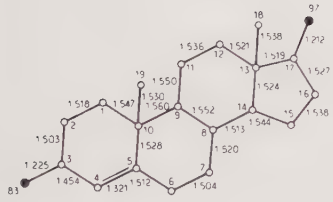


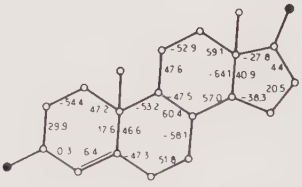




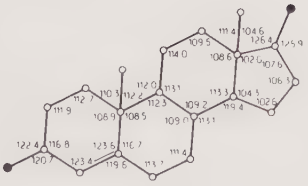
TESTOSTERONE

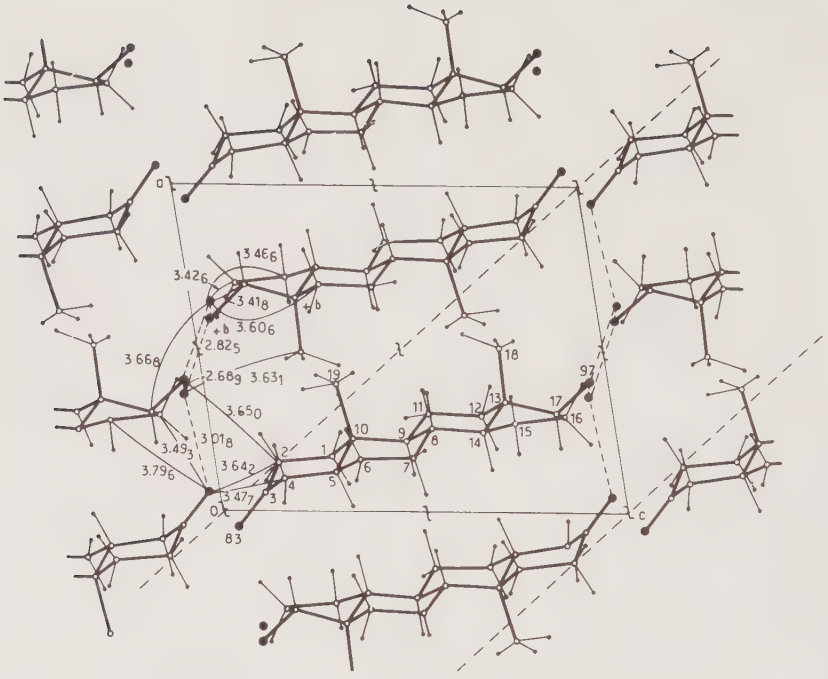




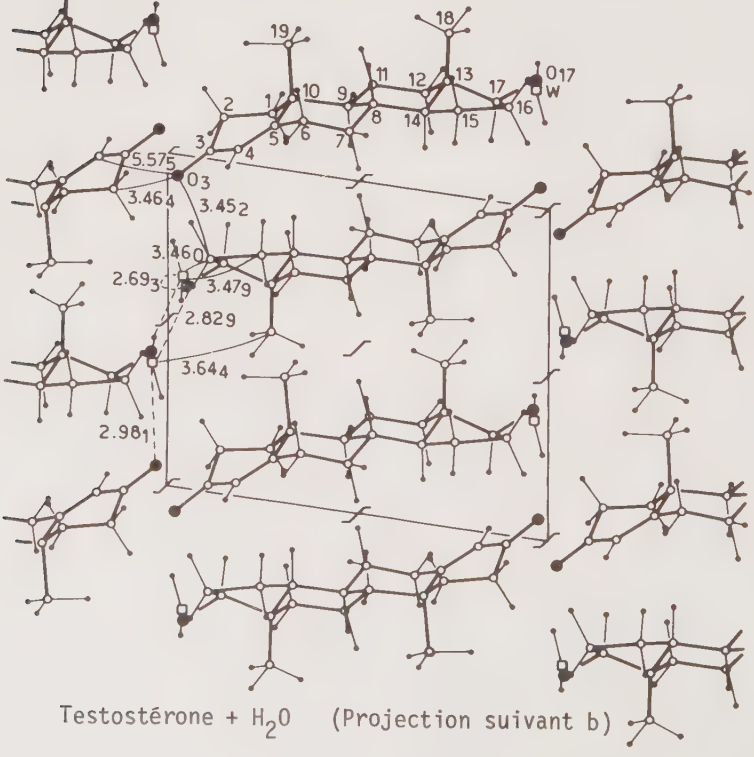


ANDROSTENEDIONE

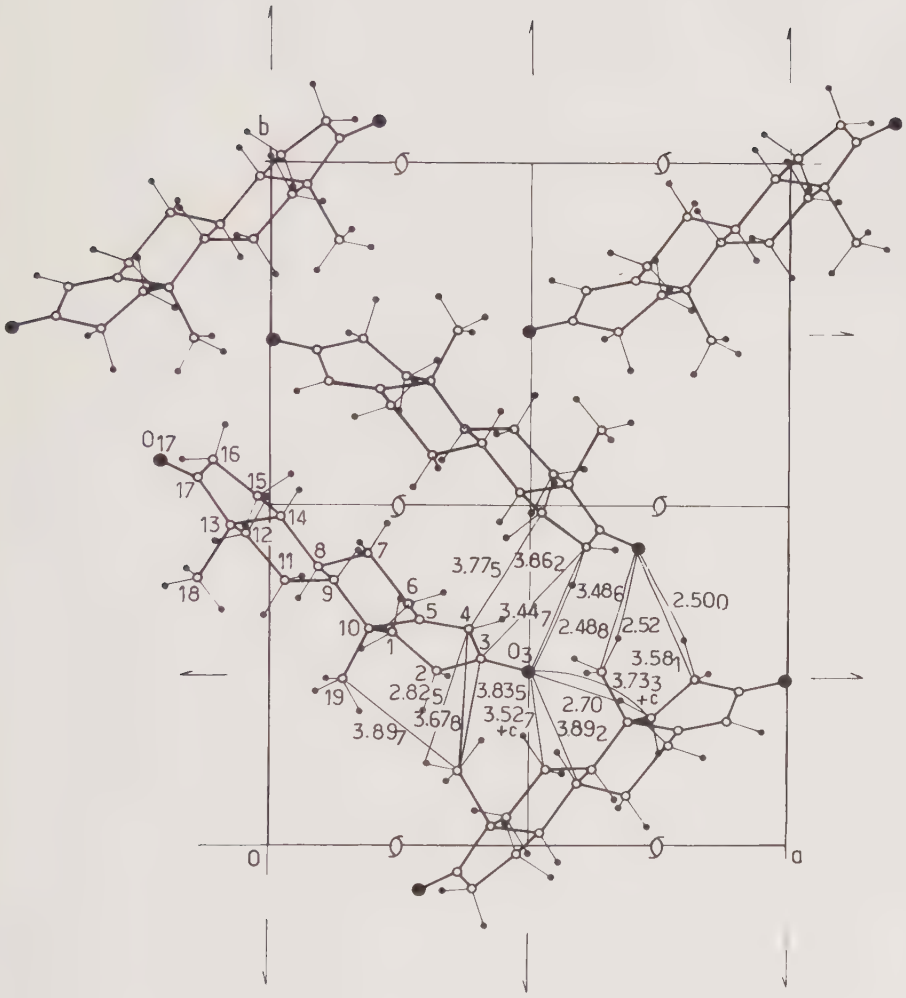




Androstanolone + H_2O (Projection suivant b)



Testostérone + H_2O (Projection suivant b)



Androstènedione (Projection suivant c)

D E S ANHYDRE

B.BUSETTA et M.HOSPITAL C.R.Acad.Sc.Paris 268 p.2011 (1969)
E.SMILEY et M.G.ROSSMANN Chem.Comm. p.198 (1969)
C.M.WEEKS, A.COOPER et D.A. NORTON - Acta Cryst. B.26 p.429 (1970)

Nous indiquons ici nos résultats. Les résultats de C.M.WEEKS sont indiqués entre parenthèses.

P b c a Z = 4 a = 18,941 ± 0,004 Å
b = 14,933 ± 0,004 V = 1497,29 Å^3
c = 5,296 ± 0,001

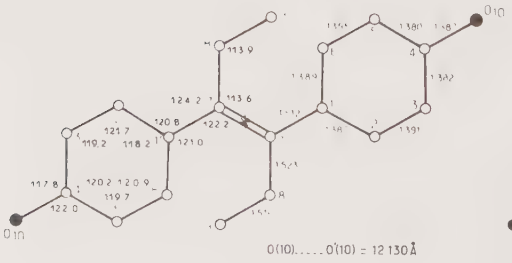
R_{final} = 0,068 (0,091) Détermination par addition symbolique des signes (par analyse d'une fonction de Patterson tridimensionnelle).

CYCLADIENE ANHYDRE

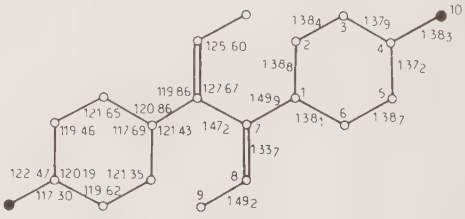
J.M.FORNIES-MARQUINA, B.BUSETTA, C.COURSEILLE, M.HOSPITAL
Acta Cryst. (1972) B 28 p.655

P b c a Z = 4 a = 18,969 Å
b = 14,296 V = 1458,95 Å^3
c = 5,380

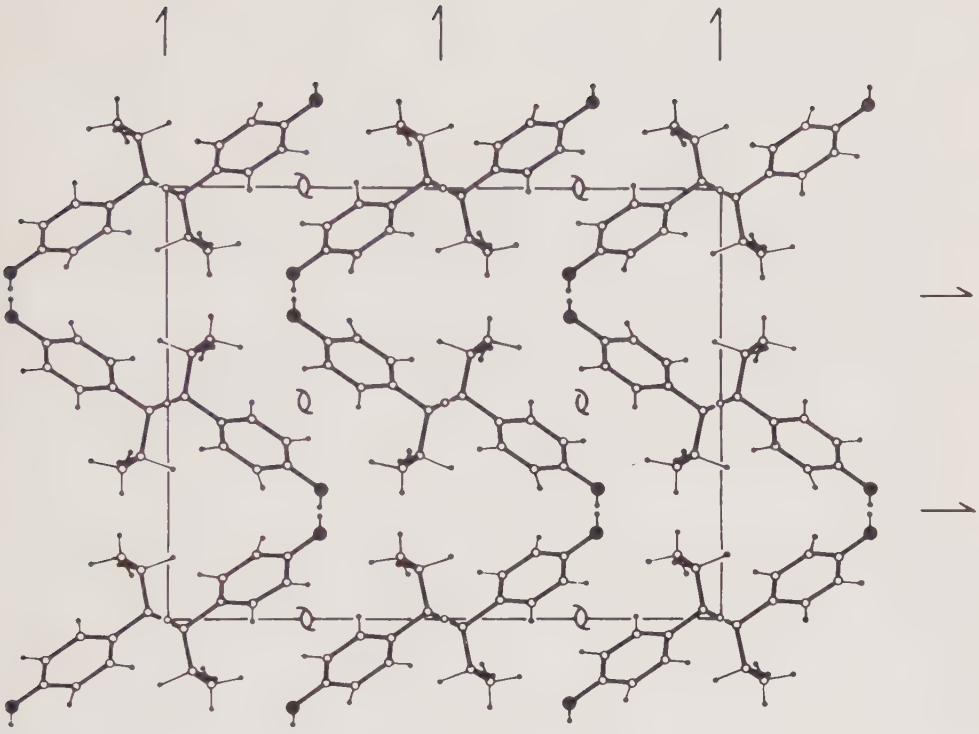
R_{final} = 0,059 Détermination par addition symbolique des signes.



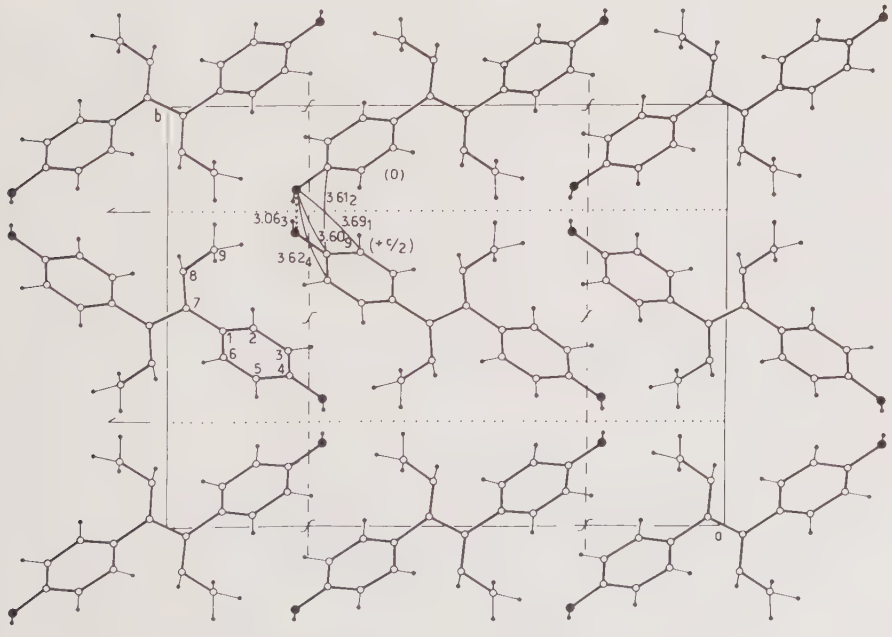
D.E.S



CYCLADIENE



DES anhydre (Projection suivant c)



Cycladiene (Projection suivant c)

DES + METHANOL + H_2O

B.BUSETTA et M.HOSPITAL
C.R.Acad.Sc.Paris t.269 (1969) p.1521

$P\bar{1}$	Z = 2	a = 9,282	α = 73°45	
		b = 13,657	β = 79°92	V = 882,6 $\mathring{A}^3$
		c = 7,676	γ = 100°92	

R_{final} = 0,069 Détermination par analyse de la fonction de Patterson.

DES + DIMETHYL SULFOXYDE (DMSO)

G.COMBERTON, Mme F.LEROY - C.R.Acad.Sc.Paris (1971) t.273 p.1160

$P\bar{1}$	Z = 2	a = 9,301 $\mathring{A}$	α = 105°80	
		b = 13,664	β = 79°61	V = 925,6 $\mathring{A}^3$
		c = 7,883	γ = 88°83	

R_{final} = 0,062 Détermination par analyse de la fonction de Patterson (DES) puis minimisation directe du facteur résiduel (DMSO).

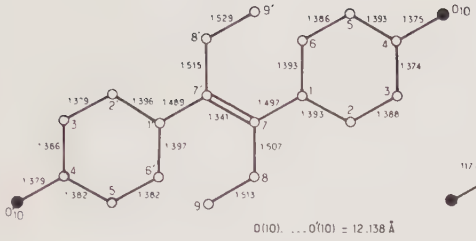
DES + ETHANOL

B.BUSETTA, Mme F.LEROY, C.COURSEILLE, M.HOSPITAL
C.R.Acad.Sc.Paris t.272 (1971) p.1304

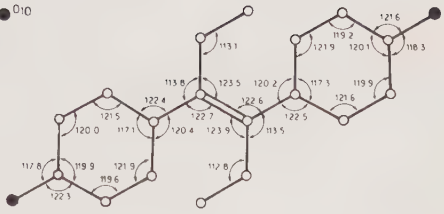
$P\bar{1}$	Z = 2	a = 9,294 $\mathring{A}$	α = 71,95°	
		b = 13,841	β = 80,26°	V = 925,6 $\mathring{A}^3$
		c = 7,713	γ = 98,60°	

R_{final} = 0,060 Détermination par analyse de la fonction de Patterson.

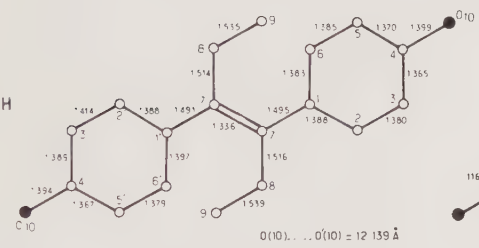
DES + CH_3OH + H_2O



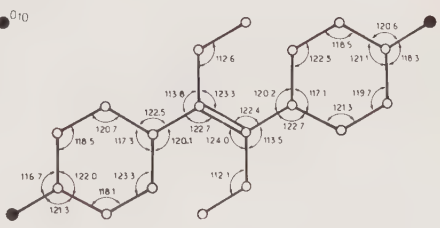
O(10)... O'(10) = 12.138 Å



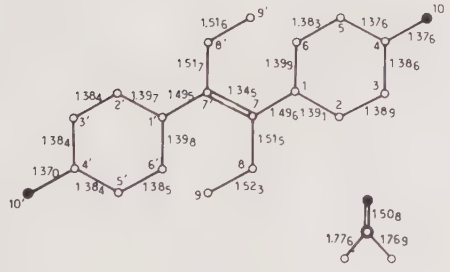
DES + C_2H_5OH

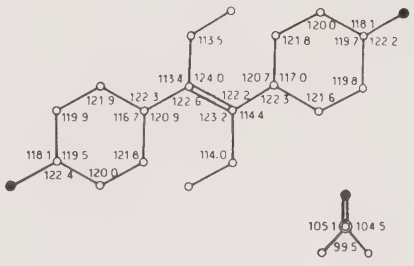


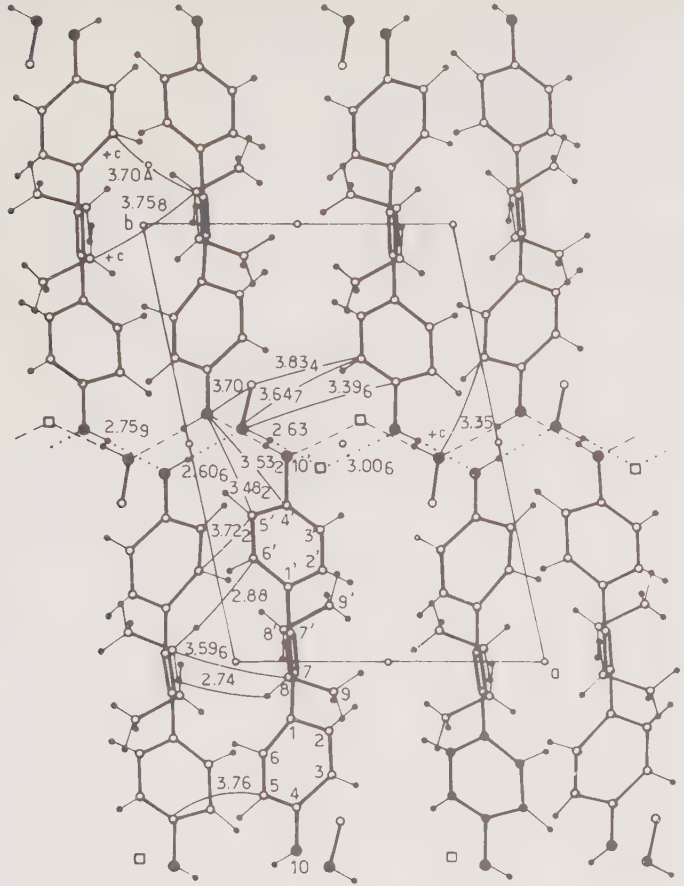
O(10)... O'(10) = 12.139 Å



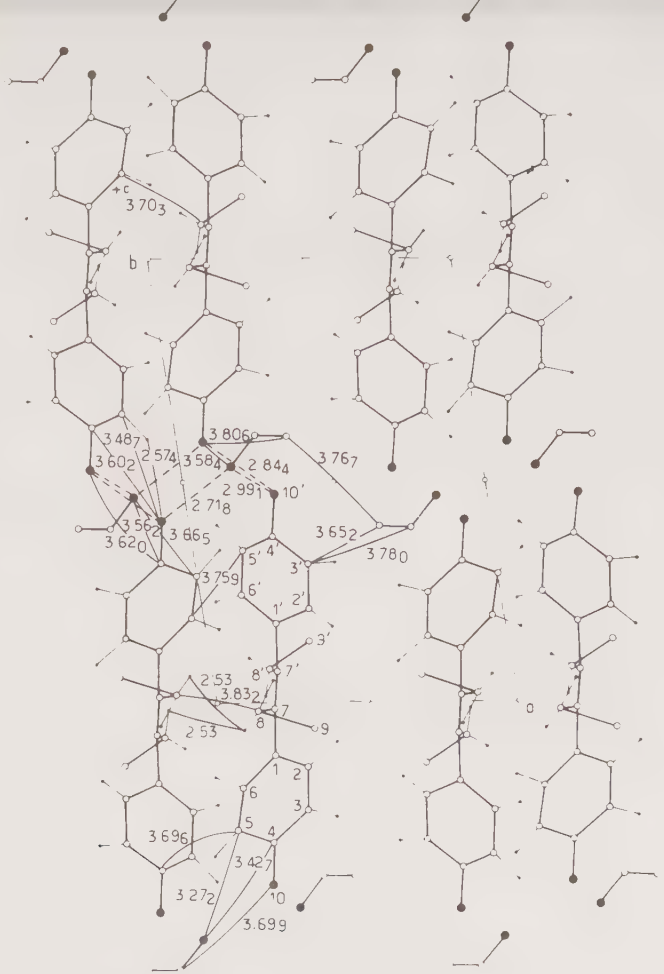
DES + DMSO



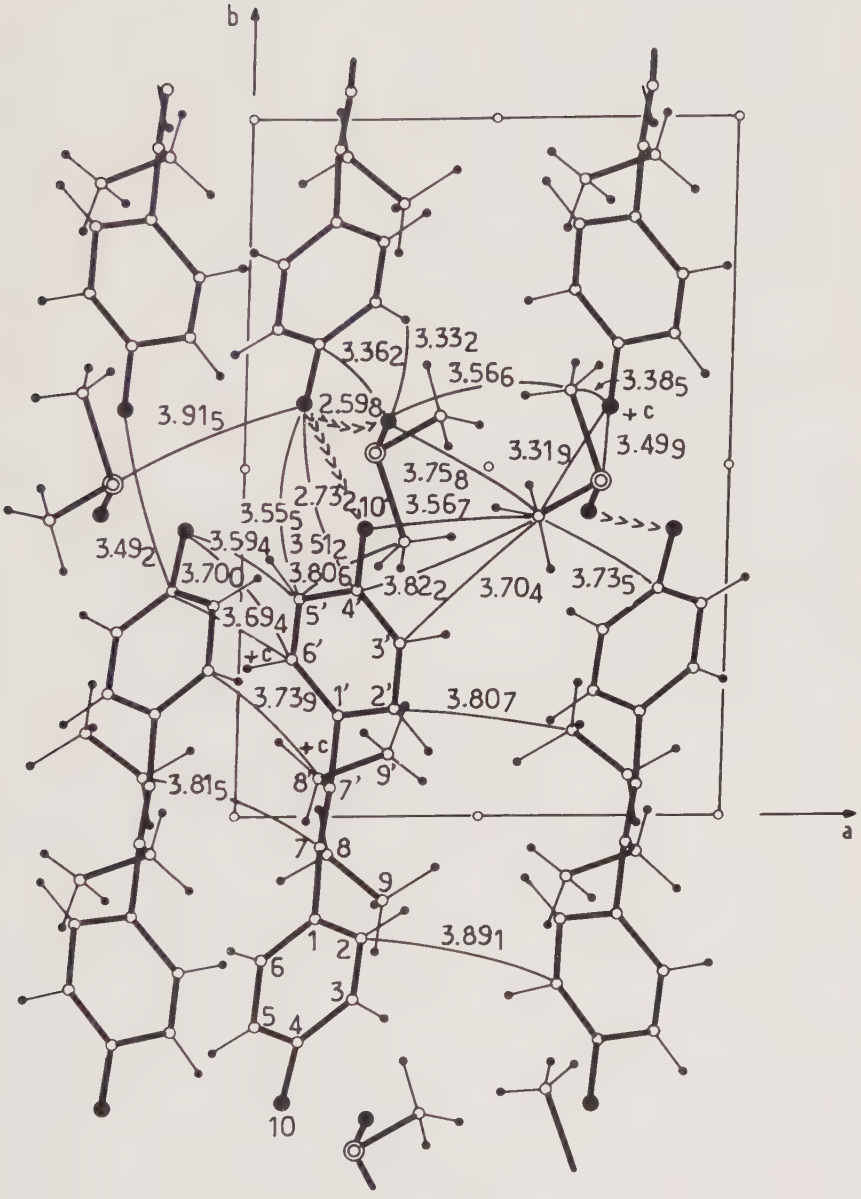




DES + CH_3OH + H_2O (Projection oblique parallèle à c)



DES + C_2H_5OH (Projection oblique parallèle à c)



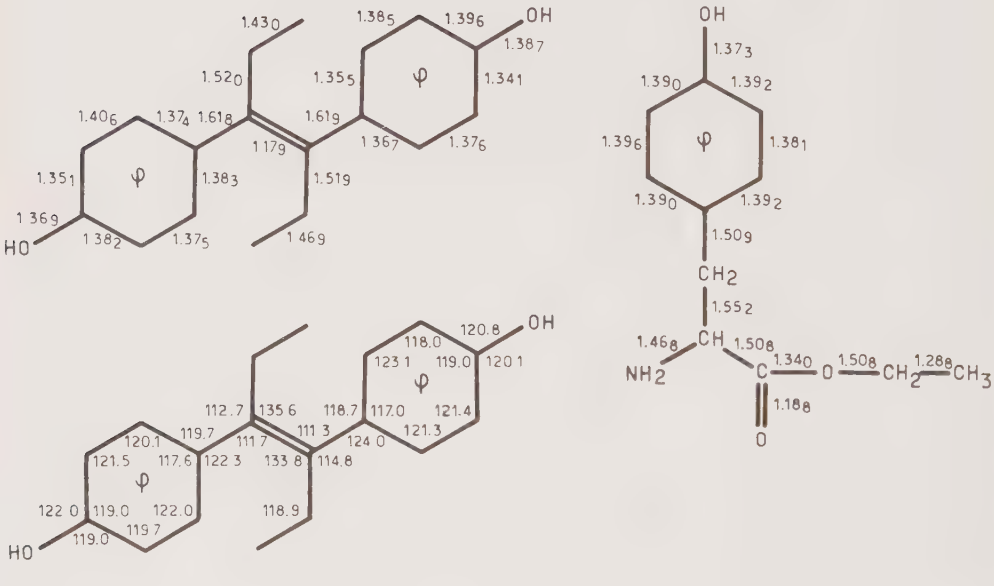
DES + DMSO (Projection oblique parallèle à c)

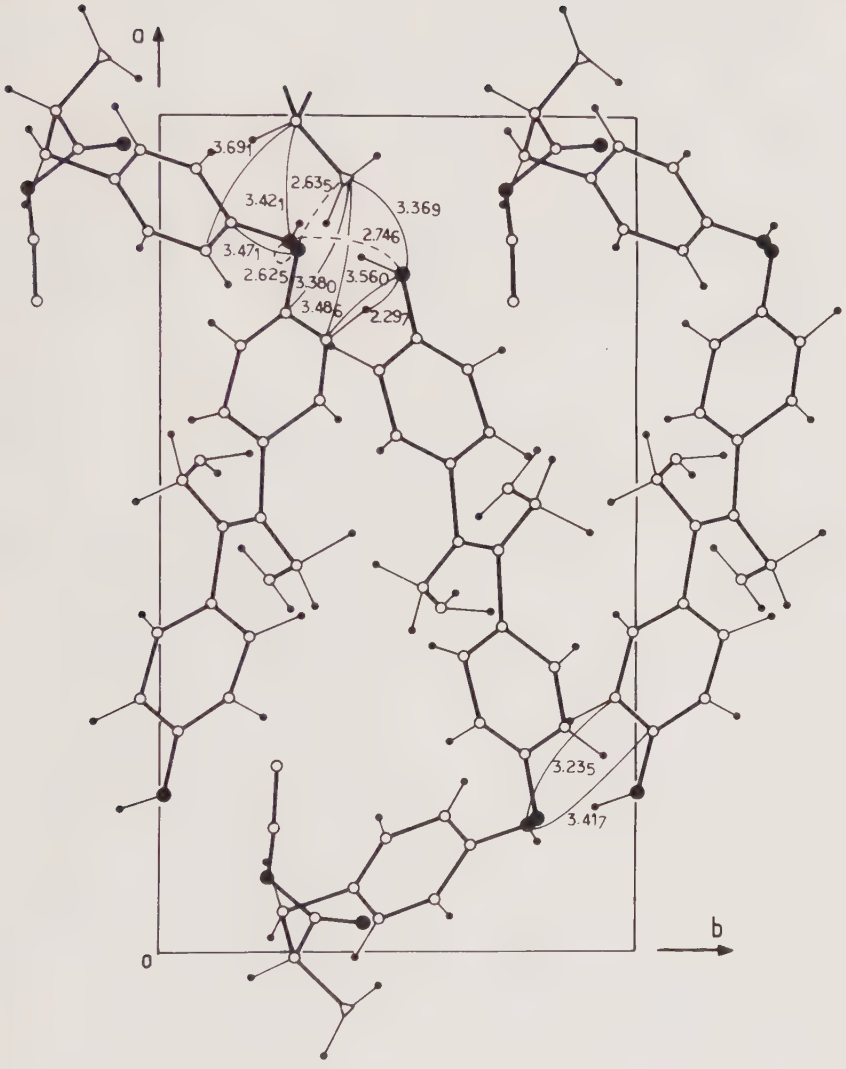
D E S + ℓ TYROSINE ETHYL ESTER

(non publiée)

$P2_1$ Z = 2 a = 17,498 Å b = 9,839 Å c = 8,213 β = 104°05'

R final = 0,073 Détermination par utilisation simultanée de l'addition symbolique et de la multisolution.





DES + ℓ Tyrosine éthyl ester (Projection oblique parallèle à c)

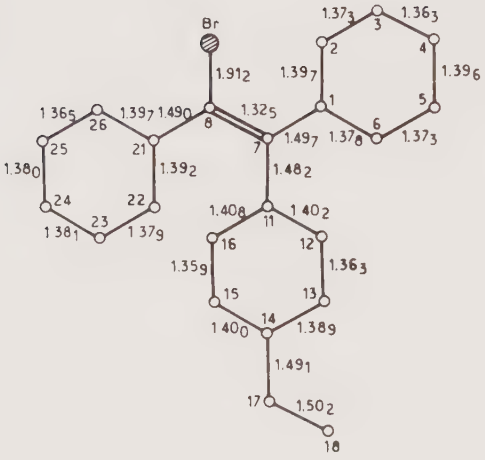
BROPAROESTROL ou 1 BROMO-2 (p ETHYL PHENYL)-1,2 DIPHENYLE ETHYLENE

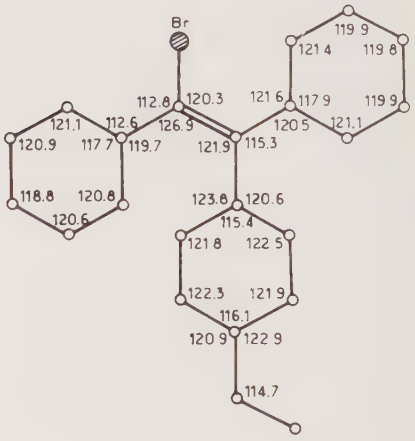
J.M.FORNIES-MARQUINA, C.COURSEILLE, B.BUSETTA, M.HOSPITAL

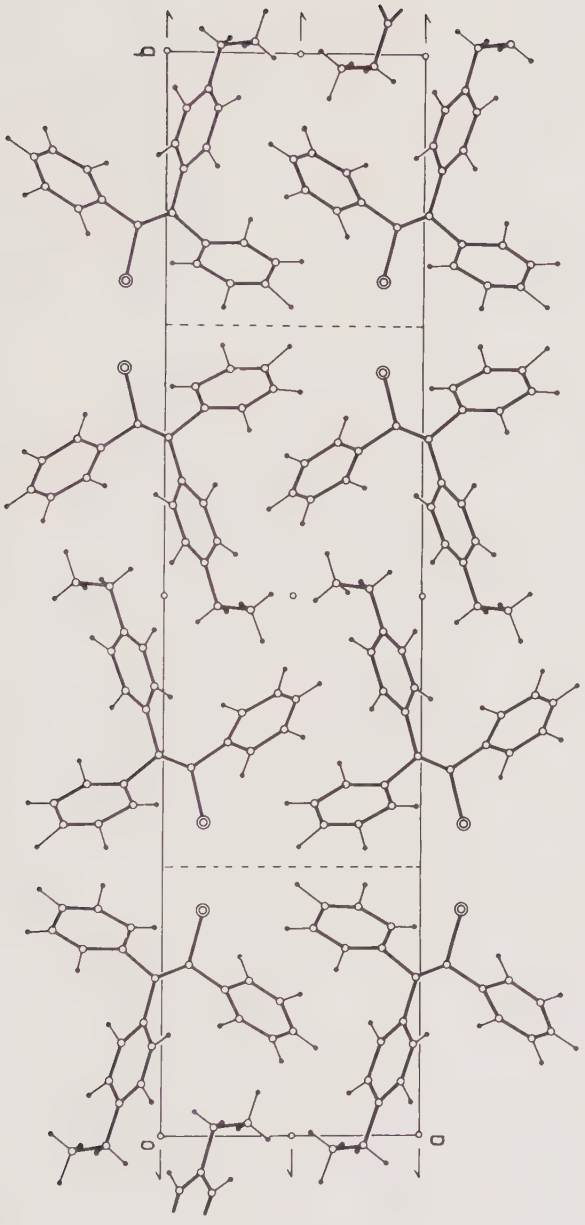
Cryst.Struct.Comm. (1972) 1 p.261

$P2_1/c$ Z = 4 a = 8,574 b = 35,767 Å c = 5,767 β = 86°1

R final = 0,084 Détermination par la méthode de l'atome lourd.







Broparoestrol (Projection suivant c)

DEUXIEME SECTION

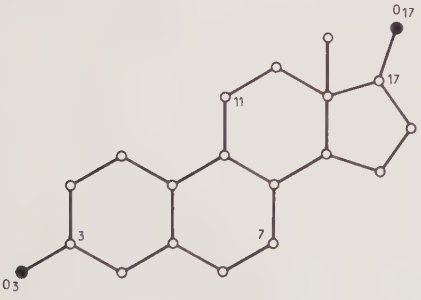
ETUDE DE LA RELATION STRUCTURE-ACTIVITE DANS LE CAS DES HORMONES OESTROGENES

PREMIERE PARTIE

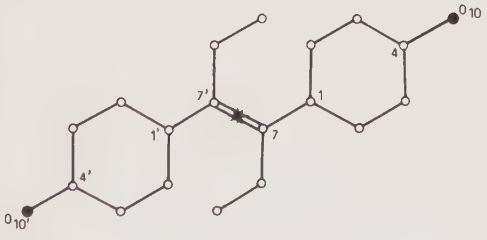
OESTRADIOL ET DIETHYLSTILBOESTROL

INTRODUCTION

Dans les années qui suivirent l'isolement de l'oestrone, cristalliséepar Doisy, l'étude de la relation structure-activité des composés oestrogènes fut rapidement développée grâce à l'utilisation intensive de synthèses chimiques de produits forts différents, dont le choix n'était basé que sur l'analogie de leurs représentations bidimensionnelles. Le plus grand succès de cette époque fut la découverte par Dodds en 1938 du diéthylstilboestrol, dont la formule plane, "arrangée" de façon convenable, pouvait présenter une certaine ressemblance avec celle de l'oestradiol:: les 2 groupements éthyls portés par la chaîne centrale, esquissant les cycles B et C du stéroïde,permirent à Dodds de formuler une théorie des cycles potentiels.



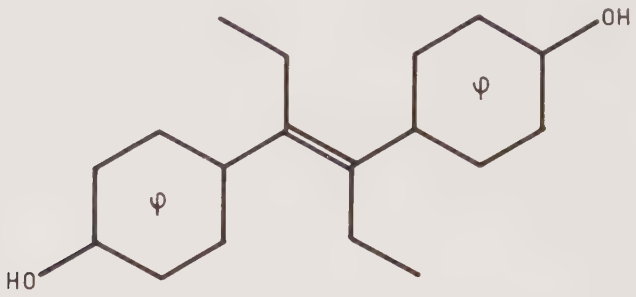
OESTRADIOL



D.E.S

Cependant, bien vite, la simplicité d'un tel raisonnement allait être mise en évidence : Koch montrait en effet, en tenant compte d'une

part des propriétés de résonance des différents modèles structuraux envisageables pour le DES, et d'autre part de sa conformation centrosymétrique, qu'il était nécessaire que la chaîne hexènique possède une configuration trans-trans, les 2 noyaux benzéniques faisant alors un angle d'environ 50° avec le plan de cette chaîne (14).



De plus apparurent rapidement des composés dont l'activité semblait totalement inexplicable : octofolline, triphényl éthylènes,... . Si bien que dans ses "Mises au point" sur ce problème (11), J.Jacques reconnaissait en 1949 qu'il manquait à la théorie spatiale de l'activité des oestrogènes "l'appui décisif que pourrait lui fournir les modèles moléculaires tirés de l'interprétation complète des spectres de rayons X". Il a fallu une vingtaine d'années pour que la cristallographie ait fait suffisamment de progrès pour permettre une telle étude.

Mais l'obtention des structures cristallines ne peut, à elle seule, fournir une explication satisfaisante de l'activité oestrogène. Pour que les résultats apportés par le cristallographe prennent leur valeur réelle il est nécessaire que les biochimistes puissent les utiliser. Il fallait donc également que la connaissance biochimique de l'activité oestrogène ait atteint un développement suffisant pour rendre efficace l'étude de la relation structure-activité au moyen de la cristallographie.

* *
*

Au début l'activité d'un composé oestrogène était définie à partir de l'observation statistique "in vivo" de ses effets primaires ou même secondaires (nombre d'oestrus provoqués, augmentation du poids de l'utérus,...). Ce genre d'étude présente des graves inconvénients :

- il est impossible de comparer des résultats qui correspondent généralement à des manifestations différentes.

- les résultats obtenus dépendent du choix des individus dans l'échantillon considéré.

- et surtout, on ne peut jamais être sûr que le corps testé représente réellement l'élément actif ou s'il ne constitue qu'un simple précurseur : on néglige totalement l'action des différents métabolismes.

Or il est prouvé maintenant (2) qu'il existe au niveau des organes cibles des structures hautement différenciées capables de reconnaître le signal "hormone", de le recevoir et de provoquer les réactions cellulaires : pour produire leur effet les hormones oestrogènes vont entrer d'une façon ou d'une autre en contact direct avec des macromolécules, les récepteurs.

Cette association oestrogène-récepteur va présenter une spécificité quantitative :

- tous les oestrogènes, quelle que soit leur forme chimique, se fixent sur le récepteur; par contre les composés non oestrogènes, même s'ils possèdent une structure stéroïdique, ne se fixent pas. De plus, le nombre de sites de fixation est limité contrairement à ce qui se produit pour les associations faibles non spécifiques (par exemple dans l'association stéroïde-albumines) pour lesquelles le nombre de sites est illimité.

- en dehors des organes cibles, il n'existe pas de macromolécules capables de fixer sélectivement les composés actifs.

Cette spécificité du récepteur pour les oestrogènes ne peut pas être assurée par un polysaccharide qui a généralement une structure peu diversifiée, ni par un lipide dont le seul caractère hydrophobe ne peut expliquer la sélectivité, ni enfin par un acide nucléique qui fixe difficilement les stéroïdes :

Seule une protéïne peut offrir à l'action hormonale des possibilités mécaniques suffisantes.

L'identification du récepteur utérin à une protéïne a d'autre part été prouvée par le fait qu'il est détruit :

- par les protéases (pronase, trypsine) alors que la ribonucléase et la désoxyribonucléase sont sans action.

- par un chauffage à 60° pendant 1 heure .

Il a été également montré qu'au niveau cellulaire c'est l'oestrogène qui agit et non pas un de ses métabolites comme c'est le cas de la testostérone dans la série des androgènes.

On peut considérer que cet oestrogène va agir en deux temps :

- d'une part une période de fixation sur le récepteur protéinique qui va subir un certain nombre de transformations. Il est peu probable que l'hormone se fixe au niveau du site actif, mais sa fixation en un endroit différent va modifier la réactivité de ce site : Tomkins a, par exemple, montré (18) que l'oestradiol était capable de transformer l'action de la glutamodéshydrogénase en une action alanine déshydrogénase. Nous dirons que l'hormone modifie "allostériquement" la protéïne réceptrice.

- cette transformation du récepteur change alors l'équilibre qui existait dans le milieu cellulaire et entraine ce que l'on désignera sous le nom "d'action oestrogène". Ce dernier point ne peut être compris qu'une fois connu , de façon parfaite, les différents constituants des cellules en cause,

leurs rôles respectifs et même les interactions qui existent entre les différentes cellules se trouvant au niveau de l'organe-cible; cela correspond à un grand nombre de facteurs qui, pour le moment, ne semblent pas être de la compétence du cristallographe.

S'il était possible de réaliser d'une part la structure de la protéïne isolée et d'autre part la structure de la même protéïne ayant fixée un oestrogène on pourrait décrire le mode de fixation et surtout les modifications qui en résultent au niveau du récepteur.

En réalité le problème n'est pas aussi simple; même si l'on peut obtenir pour la protéïne une forme cristalline utilisable : du fait que l'association "protéïne-corps actif" est une réaction d'équilibre la cristallisation de ce complexe est vraisemblablement encore plus hypothétique.

De toute façon les structures de protéïnes ne fournissent qu'une précision de l'ordre de l'Angstroem; les résultats, obtenus alors, ne pourront être interprétés que si l'on possède, par ailleurs, des connaissances plus "fines" sur la conformation et sur la réactivité potentielle de l'oestrogène : une structure de protéïne peut permettre de savoir "comment" se fixe le corps actif, elle ne permettra généralement pas de savoir "pourquoi" il se fixe mieux qu'un autre oestrogène et surtout pourquoi un dérivé très voisin ne se fixe pas.

Ainsi, même si la structure de la protéïne-cible était connue, il serait nécessaire d'entreprendre une étude cristallographique sur les différents corps actifs et leurs dérivés voisins, d'une part pour connaître de façon précise leur conformation stéréochimique, d'autre part pour déterminer les types d'interactions qu'ils sont susceptibles de donner avec d'autres éléments simples (eau, alcools, acides aminés) qui se retrouveront au niveau du récepteur. Pour qu'ensuite l'étude de la structure de la protéïne-cible puisse être utilisée de façon concluante, il faudra avoir réuni le plus grand nombre possible de renseignements à l'échelon des petites molécules et avoir réalisé un nombre suffisamment important de structures de complexes. On disposera alors d'un éventail assez large de schémas possibles pour la fixation du corps actif.

Cette dernière section voudrait constituer une première approche de cette immense entreprise.

* *

*

CHAPITRE I

LES DEUX CONFORMATIONS DU DIETHYLSTILBOESTROL

I-1 INTRODUCTION

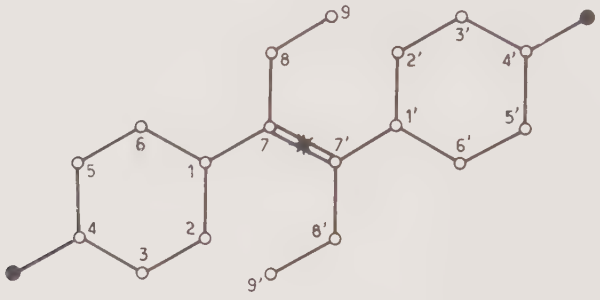
A l'époque de sa synthèse, le diéthylstilboestrol (DES) fut considéré comme une structure plane analogue au stilbène. H.P.Koch (14) fut le premier à proposer une structure non plane où les cycles benzéniques feraient un angle de l'ordre de 50° avec le plan de la chaîne aliphatique centrale, celle-ci présentant alors une configuration trans-trans plutôt que la configuration S-cis admise par Dodds(7). Cependant Koch continuait d'admettre pour le DES une seule conformation possible, centrosymétrique, et les différentes analyses conformationnelles et études physico-chimiques entreprises depuis n'ont jamais démontré qu'il pouvait en être autrement.

L'étude de ce composé par voie cristallographique était une entreprise bien tentante et plusieurs chercheurs ont entrepris l'étude de différentes formes cristallines comme en témoignent les nombreuses mailles publiées depuis cette époque. Notamment Giacomello (9) a décrit la forme triclinique obtenue dans l'alcool éthylique que l'on étudiera dans ce chapitre, mais il n'a pas pu résoudre la structure.

Brusquement ces dernières années, la résolution de la structure du DES a, semble-t-il, intéressé les cristallographes et la forme centrosymétrique a été réalisée simultanément par 3 équipes différentes, dont la nôtre (voir annexe).

Cependant nous avons été les premiers à démontrer que, sous certaines conditions, le DES pouvait présenter une conformation différente à symétrie pseudobinaire presque aussi stable que la forme centrosymétrique. L'étude de ces 2 conformations différentes du DES fait l'objet du présent chapitre.

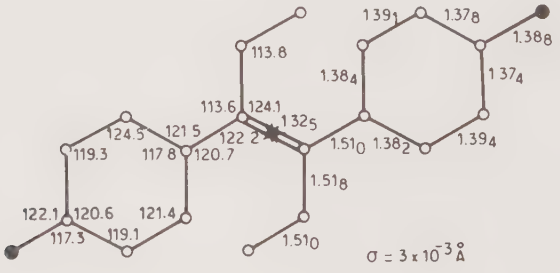
Rappelons que le diéthylstilboestrol ou α α' diéthyl 4-4' dihydroxystilbène correspond à la formule chimique :



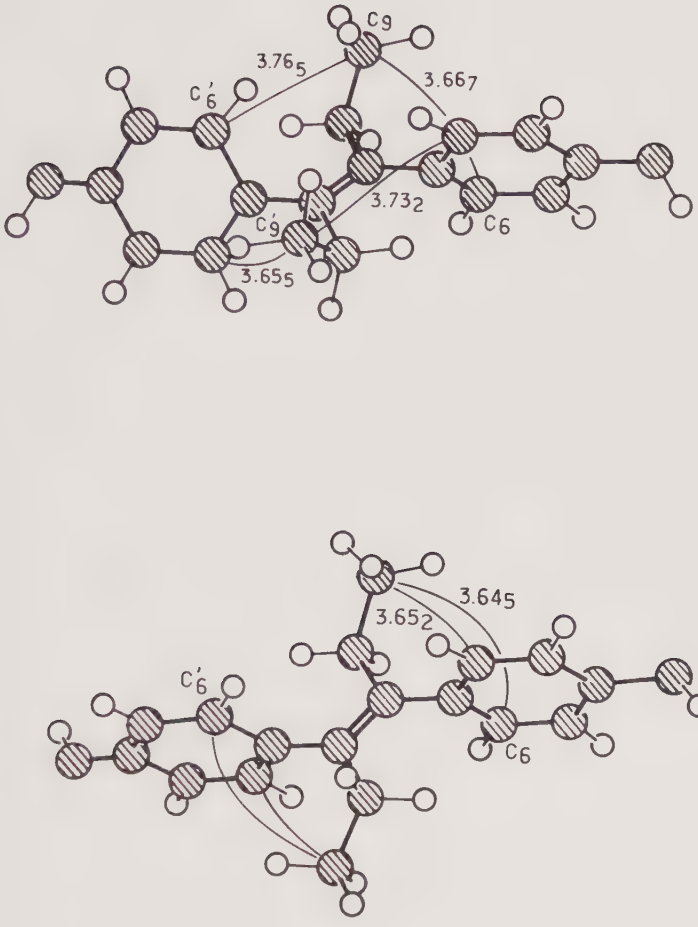
Les positions 7 et 7' sont également désignées par α et α' . L'aspect général des deux conformations du DES peut être représenté par la figure de la page suivante .

I-2 LE DIETHYLSTILBOESTROL CENTROSYMETRIQUE ($DES_{(\bar{1})}$)

Nous disposons pour cette conformation, de deux études différentes de la même forme cristalline : nos résultats personnels sont en bon accord avec ceux obtenus par C.M.Weeks (21). Nous avons utilisé ces 2 séries de mesure pour définir une structure moyenne.



Nous discuterons les différentes distances interatomiques à la fin de ce paragraphe, remarquons seulement pour l'instant que la distance C_8-C_9 anormalement courte, doit vraisemblablement être imputée à la forte agitation de l'atome C_9.



Aspect général des molécules de DES$_{"2"}$ (en haut) et DES$_{\bar{1}}$

Les angles de valence du cycle benzénique sont conformes à la valeur théorique de 120°. On notera toutefois la fermeture de 2°2 de l'angle de valence C_2-C_1-C_6 qui se répercute de façon symétrique sur les angles de valence des atomes adjacents C_6 et C_2 par une ouverture de 1°4. Elle correspond également à une localisation de l'atome C_1 légèrement extérieure (0,02 Å) au plan moyen du cycle phénolique.

Les angles de valence intéressant les groupements éthyles, diffèrent également de façon importante des valeurs théoriques. L'angle de valence C_1 - C_7 - C_8 est nettement plus fermé (113°6 au lieu de 120° pour un carbone sp^2), au contraire l'angle C_7 - C_8 - C_9 est plus ouvert (113°8 au lieu de 109°).

Nous retrouverons les mêmes déformations dans la structure à symétrie pseudo-binaire dans laquelle les atomes terminaux occupent des positions différentes par rapport au reste de la molécule.

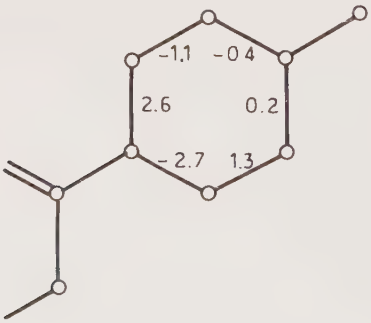
C.Courseille a également observé des écarts similaires pour les structures de la stilbamidine et de l'hydroxystilbamidine (6) où les atomes C_7 et $C_{7'}$ ne sont pas substitués.

Nous pensons donc que ces déformations ne dépendent absolument pas des groupements éthyles et notamment de leur orientation, mais plutôt d'interactions entre la portion plane de la chaîne aliphatique centrale avec les cycles phényles. Dans le cas présent, nous aurons essentiellement trois types d'interactions :

interactions $H_{2'}$-C_8 (d = 3,02 Å) qui ouvre l'angle de valence $C_{7'}$,C_7-C_8

et interactions $C_{7'}$-H_2 et C_8-H_6 qui ferme l'angle de valence C_2-C_1-C_6

Enfin, pour en finir avec les angles de valence, nous remarquerons que les angles relatifs à la liaison C-O phénolique sont nettement dissymétriques. Le phénomène de répulsion entre atomes d'oxygène liés par liaisons hydrogène que nous avons déjà décrit pour les structures stéroïdiques (5) permet d'expliquer l'ouverture de l'angle de valence C_3-C_4-O dans ce cas.

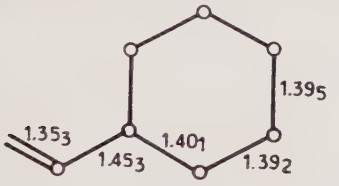


Les valeurs obtenues pour les angles dièdres du cycle phényle traduisent la localisation de l'atome C_1 en dehors de ce cycle.

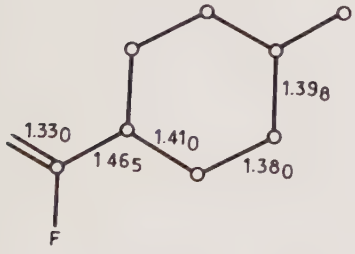
L'ensemble des atomes C_1 , C_7 , C_8 , $C_{1'}$, $C_{7'}$, $C_{8'}$ est plan; l'écart au plan moyen est au plus égal à 3 x 10^{-3} Å. Les atomes terminaux des groupements éthyles sont situés par contre nettement à l'extérieur de ce plan ($1,16_5$ Å). L'angle dièdre autour de la liaison C_7-C_8 est égal à 57°1 en valeur absolue. Du fait de la présence du centre de symétrie, la chaîne aliphatique centrale a une conformation trans-trans. L'orientation des hydrogènes de l'atome C_9 par rapport à ceux de l'atome C_8 est conforme à ce que l'on observe dans l'éthane (conformation étoilée).

Les 2 cycles phényliques sont parallèles mais ils forment avec le plan moyen de la chaîne centrale un angle de 62°8 car l'encombrement stérique des éthyles empêche une structure absolument plane.
(Rappelons que Koch avait prévu théoriquement un angle de 50°).

Dans la molécule de diéthylstilboestrol, on pourrait s'attendre à un phénomène de conjugaison entre les électrons π des cycles et ceux de la double liaison centrale. Cette conjugaison est maximale si l'ensemble de la molécule est plane : c'est ce qui se produit dans le cas du stilbène où il n'y a pas d'empêchements stériques. Les différents électrons π sont fortement délocalisés : la liaison C_1-C_7 prenant un caractère de double liaison partielle se raccourcit, au contraire la liaison C_7-$C_{7'}$ a tendance à être plus longue (16). A l'intérieur du cycle benzénique les liaisons C_2-C_3 et C_5-C_6 sont légèrement plus courtes que les liaisons adjacentes.



Mais dès que les cycles tournent et à mesure que l'angle de rotation augmente, la délocalisation électronique diminue. Dans la structure du 4-4' diméthyl α α' difluorostilbène Tchetkina et Golder (17) observe une rotation d'environ 26° autour de C_1-C_7.
Cette liaison est plus allongée, alors que la liaison C_7-$C_{7'}$ s'est raccourcie.



Dans notre cas, où l'angle de rotation est plus important, la diminution de la conjugaison est encore plus marquée : la liaison C_7-$C_{7'}$, égale à $1,32_5$ Å, a pratiquement un caractère de double liaison pure, alors que la liaison C_1-C_7, maintenant égale à $1,51_0$ Å, correspond à la distance théorique ($1,50_5$ Å) que Bartell a déterminée pour une liaison sp^3 - sp^2(1b) en l'absence de tout phénomène de conjugaison. De ce fait, si la molécule de DES est le siège d'un tel phénomène, il est assurément très faible et de peu d'importance énergétique.

Enfin, dans cette forme de DES, on observe au niveau du cycle benzénique des liaisons interatomiques significativement différentes (C_2-C_3

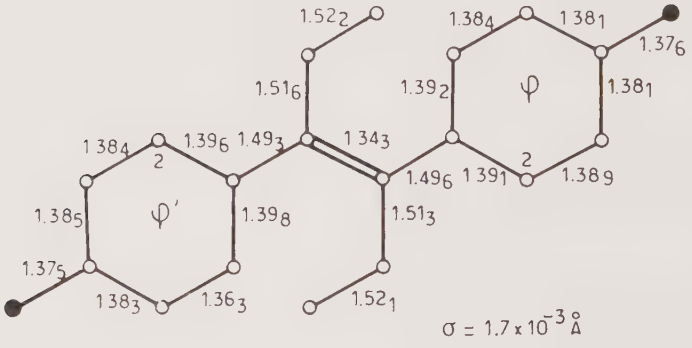
et C_5-C_6 sont plus longues que les liaisons adjacentes) ce phénomène ne se retrouve pas dans les structures pseudobinaires que nous allons décrire maintenant. Peut-être est-il lié à un état électronique différent dans les 2 cas ?

I-3 DIETHYLSTILBOESTROL A SYMETRIE PSEUDOBINAIRE (DES"2")

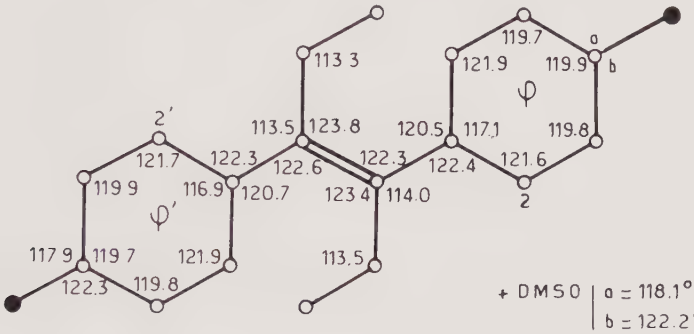
L'existence d'une telle conformation fut longtemps mise en doute, principalement pour des raisons d'encombrement stérique (19). Nous l'avons mise en évidence pour la première fois dans un cristal obtenu à partir d'une solution de diéthylstilboestrol dans l'alcool méthylique, mais depuis de nombreux autres solvates du DES (avec l'éthanol, le diméthylsulfoxyde,...) ont présenté la même conformation.

En réalité, nous devons tout de suite faire remarquer que cette conformation ne correspond pas à une molécule symétrique, et nous verrons au contraire, qu'au niveau des atomes d'oxygène, elle est, dans le cristal, fondamentalement dissymétrique. Cependant, par commodité, nous continuerons à l'appeler DES"2" désignation qui, dans notre esprit, traduit seulement l'aspect général de la molécule.

Les figures ci-dessous représentent les distances et angles de valence de la molécule moyenne construite à partir des 3 structures cristallines que nous avons étudiées.



Il ne semble pas y avoir de variations importantes de la conjugaison par rapport à la forme centrosymétrique. Les liaisons C_1-C_7 et $C_{1'}$-$C_{7'}$ ne diffèrent pas énormément des valeurs précédemment observées ; elles sont, cependant, légèrement plus courtes (respectivement 1,49$_6$ Å et 1,49$_3$ Å au lieu de 1,51$_0$ Å dans le DES$_{\bar{I}}$). Par la suite, si effectivement il existe une légère différence d'énergie de conjugaison entre les 2 conformations, les valeurs observées pour les longueurs de liaison du DES et du DES$_{\bar{I}}$ inciteraient à attribuer à la forme pseudo-binaire la conjugaison la plus forte.



Au point de vue angles de valence, il n'y a également pas de différence significative avec la structure centrosymétrique. Par la suite, nous voyons que les interactions qui agissent sur les angles de valence C_2-C_1-C_6 et C_1-C_7-C_8 doivent avoir la même amplitude dans les 2 conformations. Il est alors vraisemblable qu'elles font intervenir uniquement les atomes C_7 $C_{7'}$ C_8 $C_{8'}$ de la chaîne centrale.

Pour les angles de valence relatifs à la liaison C_4-O_{10} du stilboestrol, les déviations se produisent en sens inverse dans les structures DES + CH_3OH et DES + C_2H_5OH d'une part et dans le DES+DMSO d'autre part. Si on regarde les projections représentées en annexe, on peut voir que le phénomène de répulsion apparente des atomes d'oxygène liés par liaison hydrogène dont nous avons déjà parlé rend parfaitement compte des écarts observés.

Dans les deux solvates alcooliques, l'atome d'oxygène O_{10} reçoit un doublet électronique de l'oxygène de la molécule d'alcool; nous avons une liaison O_{10} - H ... $O_{alcoolique}$ qui va ouvrir l'angle de valence C_5-C_4-O_{10} au détriment de l'angle C_3-C_4-O_{10}.

Nous observons effectivement :

- dans la structure DES + CH_3OH + H_2O : C_5-C_4-O_{10} = 121°6 et C_3-C_4-O_{10}= 118°3
- sans la structure DES + C_2H_5OH : C_5-C_4-O_{10} = 120°6 et C_3-C_4-O_{10}= 118°3

(Cependant dans cette structure nous n'avons pas pu placer les atomes d'hydrogène engagés dans des liaisons, mais nous avons supposé que l'enchainement donneur-accepteur était réalisé de la même façon que dans le DES méthanol et notamment qu'entre les oxygènes du DES s'établissait une liaison hydrogène du type $O_{10'}$ - H ... O_{10}).

Dans la structure du DES + DMSO la liaison hydrogène intéressant l'atome O_{10} implique au contraire une ouverture de l'angle de valence C_3-C_4-O_{10}. Nous n'avons pas, ici non plus, pu placer les atomes d'hydrogène mais on connait exactement l'enchainement donneur-accepteur car l'oxygène du diméthylsulfoxyde ne peut être qu'un donneur d'électrons. Les angles de valence effectivement observés dans cette structure sont :

$$C_3\text{-}C_4\text{-}O_{10} = 122°2 \quad \text{et} \quad C_5\text{-}C_4\text{-}O_{10} = 118°1$$

Si les distances et angles de valence observés dans les 2 formes de diéthylstilboestrol $DES_{\bar{1}}$ et $DES_{"2"}$ sont peu différentes il n'en est pas de même de la géométrie de ces molécules :

- dans la forme pseudo binaire les deux cycles phényles sont croisés (ils forment entre eux un angle respectivement égal à 59°, 61°, 70°, dans les structures avec CH_3OH, C_2H_5OH et le diméthylsulfoxyde), cependant ils continuent à former avec le plan moyen de la chaîne aliphatique centrale des angles à peu près égaux (57° et 63° environ dans les 2 solvates avec CH_3OH et C_2H_5OH).

- les deux éthyles qui étaient en position trans-trans (c'est-à-dire situés de part et d'autre du plan de la chaîne centrale(C_1 C_7 $C_{7'}$ $C_{1'}$)sont maintenant en conformation S.cis (c'est-à-dire d'un même côté du plan de la chaîne centrale).

Par la suite on observe, au niveau des atomes C_9 et $C_{9'}$, des contacts intramoléculaires faisant intervenir des atomes différents des cycles phényls :

- dans le $DES_{\bar{1}}$ l'atome C_9 est en contact avec 2 atomes du cycle aromatique Φ, l'atome C_2 (3,65 Å) et l'atome C_6 (3,64 Å).

- dans le $DES_{"2"}$ les contacts se font avec des atomes de cycles différents le méthyl C_9 est lié à la fois à l'atome C_6 (3,67 Å) du cycle Φ et à l'atome $C_{2'}$ (3,76 Å) du cycle Φ'.

au contraire le méthyl $C_{9'}$ est lié à l'atome $C_{6'}$ (3,65 Å) et à l'atome C_2 (3,73 Å) (*)

(*) *Au passage, nous laissons apprécier au lecteur à quel point cette notion de double contact est éloignée de la notion de cycle potentiel : nous sommes très loin de l'hypothèse de Dodds dans les 2 cas.*

Les atomes C_8 et $C_{8'}$ ne sont plus situés exactement dans le plan moyen de la chaîne centrale, comme auparavant, mais légèrement au dehors (0,04 Å). Les angles de torsion autour des liaisons C_7-C_8 et $C_{7'}C_{8'}$ sont par ailleurs peu différents (respectivement -64°4 et -61°4 dans le DES + CH_3OH ; -62°9 et-59°7 dans le DES + C_2H_5OH), les hydrogènes des atomes C_9 et $C_{9'}$ adoptent une conformation étoilée par rapport à ceux des atomes C_8 et $C_{8'}$, exactement comme dans la forme centrosymétrique. Nous pouvons donc considérer que pendant le passage de la forme $DES_{\bar{1}}$ à la forme $DES_{"2"}$ il n'y a aucune rotation du groupement terminal de l'éthyle.

I-4 ETUDE DE LA TRANSITION $DES_{\bar{1}} \rightarrow DES_{"2"}$

L'étude cristallographique a montré qu'il existait deux conformations différentes du diéthylstilboestrol : une forme centrosymétrique et une forme que nous avons qualifiée de pseudobinaire. Nous voudrions essayer d'étudier maintenant les conditions qui assurent le passage de la forme $DES_{\bar{1}}$ à la forme $DES_{"2"}$.

Les cristaux correspondant à la forme $DES_{"2"}$ sont instables ; ils ont tendance à perdre leur molécule de solvant plus ou moins rapidement. On obtient alors une poudre microcristalline de la forme orthorhombique Pbca du $DES_{\bar{1}}$. Ces observations permettent de penser que la forme centrosymétrique représente la forme stable de la molécule de diéthylstilboestrol isolée. Par contre la solvatation stabilise vraisemblablement la forme pseudobinaire. S'il en est ainsi, l'énergie nécessaire au passage de la forme $DES_{\bar{1}}$ à la forme $DES_{"2"}$ doit être faible, c'est-à-dire inférieure à l'énergie de solvatation.

Nous pouvons donner l'ordre de grandeur de cette énergie de transition , W_{tr} :

$$DES_{\bar{1}} \underset{W_{tr'}}{\overset{W_{tr}}{\rightleftharpoons}} DES_{"2"}$$

Elle peut en effet se décomposer en trois termes :

$$W_{tr} = \Delta W_{conj} + W_{tors} + W_{VdW}$$

- Nous avons vu aux paragraphes précédents que la différence d'énergie de conjugaison ΔW_{conj} entre les 2 formes est sans aucun doute très faible, mais devrait favoriser plutôt la forme pseudo-binaire.

- l'énergie de torsion W_{tors} peut s'écrire à l'aide d'un potentiel de Pitzer (15) :

$$W_{tors} = \tau\left[\, 1 + \cos 3\,(\Phi + 60°) \,\right] \qquad \tau = 0{,}99 \text{ Kcal/mole}$$

où Φ représente successivement les angles de rotation autour des

liaisons C_1-C_7 (rotation du phényl $\Phi = + 60°$) et C_7-C_8 (rotation du groupement éthyl $\Phi \simeq - 135°$) (*) .

L'énergie globale correspondante est de l'ordre de 3 à 4 kcal/mole. Cette énergie de torsion intervient d'ailleurs aussi dans le passage de la forme $DES_{\bar{1}}$ à la forme $DES_{"2"}$ que dans la transition inverse, si bien qu'avec les notations utilisées, le passage du $DES_{"2"}$ ou $DES_{\bar{1}}$ correspond à une énergie :

$$W_{tr'} = - \Delta W_{conj.} + W_{tors} + W'_{VdW}$$

- Enfin le terme W_{VdW} représente la barrière du potentiel d'interactions intramoléculaires que l'on doit franchir au cours de la transition. Pour avoir une idée de l'ordre de grandeur de ce terme, nous avons évalué les variations d'énergie potentielle de Van der Waals au cours de la rotation du cycle phényl Φ et du groupement éthyl C_8-C_9, en utilisant les fonctions potentielles semi-empirique. Nous avons choisi les potentiels définis par Bartell (1 a) qui, d'après l'étude faite par E.Giglio (10), semblent donner les meilleurs résultats dans l'évaluation des paramètres cristallins, la détermination de structures aussi bien que dans l'évaluation des chaleurs de sublimation (**).

La figure ci-dessous représente les courbes de niveau d'énergie potentielle, en fonction des 2 paramètres de rotation :

- rotation du phényl en abscisse

- rotation du groupement éthyl en ordonnée

L'énergie intramoléculaire correspondant au $DES_{\bar{1}}$ a été prise pour définir le niveau zéro ($E_{DES_{\bar{1}}} = 0$ kcal).

On peut voir qu'en l'absence de rotation du cycle phényl, le groupement éthyl peut occuper différentes positions (si $\Theta_\Phi = 0$ $-40° < \Theta_{eth} < +70°$ pour la courbe à 1,5 kcal). Par contre le cycle phényl ne peut guère tourner que de 40° si le groupement éthyl reste fixe. Pour passer de la forme $DES_{\bar{1}}$ à la forme pseudo binaire $DES_{"2"}$ ($\Theta_\Phi = + 60°$ et $\Theta_{eth} \simeq - 125°$) il est nécessaire de faire varier les deux paramètres de rotation en même temps. La barrière d'énergie potentiel qu'il faut franchir au cours de cette transition est égale à 1,2 kcal/mole.

(Nous remarquerons que les courbes d'énergie potentielle indiquent que le passage de la forme $DES_{\bar{1}}$ à une forme à cycles croisés où les groupements éthyl resteraient en configuration trans-trans ($\Theta_\Phi = + 60°$ et $\Theta_{eth} = + 60°$) semble possible. Jusqu'ici, cependant, une telle conformation n'a pas été mise en évidence dans le milieu cristallin).

(*) (**) et figure pages ci-jointes

(*) Si on se reporte aux vues en perspectives des 2 formes du DES (voir paragraphe I-1), on peut considérer que le passage d'une forme à l'autre est réalisée en maintenant fixe une partie de la molécule de DES-1 (le cycle Φ'et l'éthyl $C_{8'} - C_{9'}$) et en faisant tourner le reste de la molécule.

(**) Pour évaluer les potentiels d'interactions inter ou intramoléculaires, on admet en première approximation, dans le cas de molécules organiques, le principe de l'additivité des interactions atome-atome; le potentiel moléculaire est alors la somme des potentiels atomiques :

$$V = \sum_i \sum_j V_{ij}$$

On admet aussi que les atomes de même nature chimique interagissent de la même façon, indépendamment de la molécule qu'ils composent. Un atome peut alors être considéré comme formé d'une masse électriquement neutre qui va donner lieu aux interactions de Van der Waals et d'une charge résiduelle entrainant l'existence d'interactions électrostatiques multipolaires (dipolaires ou quadripolaires). Kitaigorodskii (13) a montré, à partir du calcul de chaleur de sublimation, que ce deuxième terme pouvait être négligé sans inconvénient. Les interactions atome-atome se réduisent donc aux seules interactions de Van der Waals, qui peuvent être décomposées en 2 éléments, une énergie de dispersion et une énergie de répulsion :

$$V_{ij}(r) = \frac{a_{ij}\, e^{-b_{ij} r_{ij}}}{r_{ij}^{d_{ij}}} - \frac{c_{ij}}{r_{ij}^{6}}$$

Si $d_{ij} = 0$ on a un potentiel de Lennard Jones (interactions hydrogène-hydrogène)

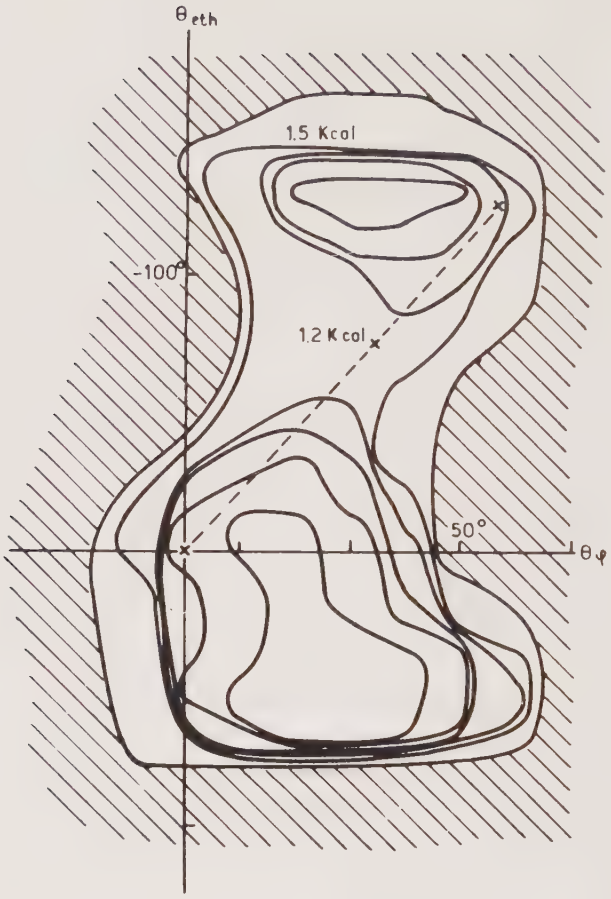
Si $d_{ij} = 12$ on a un potentiel de Buckingham (interactions carbone-carbone, et autres atomes lourds)

Les paramètres a_{ij}, b_{ij}, c_{ij} sont déterminés expérimentalement, généralement à partir d'observations en phase gazeuse. Les valeurs proposées par Bartell (1a) sont les suivantes:(l'énergie est exprimée en kcal par paire d'atomes si les longueurs sont en angstroem).

		$a \times 10^{-3}$	b	c
interactions	H-H	6,6	4,080	49,2
	C-C	301,2	0	327,2
	O-O	259,0	0	358,5

Pour les interactions mixtes C-H, O-H ou C-O, on utilise comme paramètres a et b les moyennes géométriques des termes respectifs des fonctions homonucléaires. Le paramètre c est ensuite obtenu en minimisant la fonction V(r) à la distance correspondant à la somme des rayons de Van der Waals :

		$a \times 10^{-3}$	b	b	d
interactions	C-H	44,8	2,040	125,0	6
	O-H	42,0	2,040	132,7	6
	C-O	278,7	0	342,3	12



Finalement, on peut affirmer sans trop d'erreurs que l'énergie nécessaire au passage de la forme centrosymétrique à la forme pseudobinaire est égale environ à 4ou 5 kcal/mole. Nous voudrions montrer, maintenant, que cette énergie n'est vraisemblablement pas fournie par le champ cristallin, mais qu'elle est prise plutôt sur l'énergie de solvatation. Nous pensons en effet que la forme pseudobinaire n'apparaît pas seulement au moment de la cristallisation, mais existe dès la solution.

Dès lors, elle ne devrait pas se retrouver forcément dans toutes les structures cristallines solvatées. Avec notre hypothèse, l'existence de la forme pseudobinaire dépend, en effet, de la nature du solvant : pour qu'il y ait rotation, il faut que l'énergie de solvatation correspondante soit supérieure à l'énergie de transition.

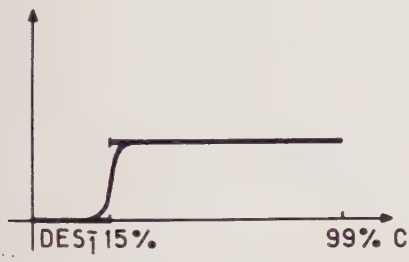
Or on peut relier l'énergie de solvatation à l'énergie des liaisons hydrogène DES-solvant observées dans le milieu cristallin. Dans la structure DES + CH_3 OH + H_2O, on observe entre la molécule de DES et les 2

molécules de solvant :

- une liaison hydrogène forte (2,60$_6$ Å) avec l'alcool méthylique
- une liaison hydrogène très faible (3, 00$_6$ Å) avec l'eau.

(la faible énergie de la liaison DES-eau est encore mise en évidence par la médiocre solubilité du DES centrosymétrique dans ce solvant).

Nous avons étudié la cristallisation du diéthylstilboestrol dans un mélange alcool méthylique-eau, en fonction de la concentration en alcool. (Nous avons, dans ce but, préparé différentes solutions hydroalcooliques de DES que nous avons fait cristalliser par refroidissement dans un tube scellé). Nous avons schématisé ci-dessous le résultat de cette expérience :



- si la concentration en alcool est faible (< à 15 %) c'est la forme centrosymétrique (Pbca) qui cristallise ; la molécule de DES ne rencontre pratiquement que des molécules d'eau, et la formation du solvate correspondant ne fournit pas suffisamment d'énergie pour faire apparaître la forme pseudobinaire.

Par contre, quand la concentration en alcool dépasse 15 % (*), il ne se forme pratiquement plus en solution que des associations DES-alcool avec une énergie de solvatation nettement plus élevée; dès cet instant, c'est la forme pseudobinaire solvatée qui cristallise.

On sait que l'aptitude d'un alcool primaire à donner des liaisons hydrogène décroît quand la chaîne hydrocarbonée augmente. Ainsi l'eau qui se lie très fortement au méthanol, s'associe moins fortement à l'éthanol et pratiquement plus au propanol. Nous avons étudié les structures du DES cristallisé dans le méthanol et l'éthanol. On observe dans les deux cas une conformation pseudobinaire; les structures cristallines de symétrie triclinique sont très voisines. Par contre, il se pourrait que les homologues supérieurs, propanol, butanol, ... ne soient pas capables d'assurer la transition DES$_{\bar{1}}$ → DES$_{"2"}$. On sait déjà que dans le propanol, le DES donne une forme cristalline différente ($P2_1/c$).

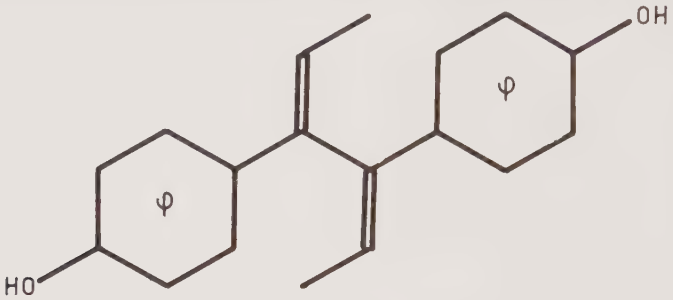
Cependant si nous pouvons admettre que la solvatation permet la transition DES$_{\bar{1}}$ → DES$_{"2"}$ quand l'énergie fournie par cette association est supérieure à un certain seuil que l'on peut fixer à 4 ou 5 kacl/mole, nous ne comprenons toujours pas très bien par quel mécanisme cet apport d'énergie entraîne le changement de conformation observée. Nous disposons vraisemblablement de trop peu d'éléments pour pouvoir conclure avec certitude.

(*) *Cette valeur pourrait d'ailleurs permettre d'estimer le rapport des constantes respectives des associations DES-eau et DES-alcool méthylique.*

I-5 STRUCTURE DU CYCLADIENE

Le diéthylstilboestrol présente 2 conformations différentes DES_I et DES_{II}. Dans la première les cycles benzéniques sont parallèles, dans la deuxième ils sont "croisés". Cependant pour passer d'une forme à l'autre on ne peut pas faire tourner le cycle phényl sans agir également sur le groupement éthyl adjacent. Nous avons vu que la torsion autour de C_1-C_7 était liée à une torsion autour de C_7-C_8.

Si on parvient à empêcher cette dernière rotation, il ne sera plus possible de faire tourner un des cycles phényls par rapport à l'autre. Pour rigidifier la liaison σ C_7-C_8 par rapport au phénomène de torsion il suffit de la transformer en une liaison π, c'est-à-dire de considérer le dérivé suivant :



Ce dérivé, qui est également un oestrogène de synthèse, est appelé cycladiène ou dièneoestrol.

Nous avons essayer de le cristalliser aussi bien dans l'hexane (c'est-à-dire dans un milieu apolaire) que dans une solution hydroalcoolique. Quelles que soient les conditions opératoires, nous obtenons une seule forme cristalline du cycladiène, non solvatée orthorhombique Pbca à molécules centrosymétriques. La rigidité de la chaîne diénique centrale empêche d'avoir plusieurs conformations différentes pour le cycladiène.

*
* *

De prime abord, la molécule de cycladiène présente les mêmes caractéristqiues que la forme centrosymétrique du diéthylstilboestrol.

Comme pour le DES et le cycle A aromatique d'un stéroïde oestrogène, le cycle phénylique admet un plan axe de symétrie contenant C_7-C_1-C_4-O_{10}; on peut le voir sur les distances interatomiques et surtout les angles de valence présentés en annexe. Ce cycle phénylique n'est pas plan et il se rap-

proche davantage du cycle d'un stéroïde oestrogène : il adopte une forme rappelant le "bateau".

Comme dans le DES, l'angle de valence C_2-C_1-C_6 est moins ouvert que l'angle théorique d'un cycle benzénique (117°7 au lieu de 120°). Par contre les angles de valence autour du carbone C_7 sp^2 sont distribués plus normalement (notamment C_1-C_7-C_8 égal à 119°9 dans le cycladiène au lieu de 113°6 dans le DES). Il est vraisemblable que les interactions entre l'atome C_8 et le cycle phényl Φ' sont moins fortes que dans le DES puisque l'atome C_8 ne porte qu'un seul hydrogène au lieu de deux précédemment, de plus, la double liaison C_7-C_8 est plus rigide.

La liaison hydrogène intéressant les atomes d'oxygène phénoliques étant identiques à celle observée dans le DES (3,06$_3$ Å contre 3,00 Å), les écarts observés sur les angles de valence C_3-C_4-O_{10} sont également identiques.

Comme dans le DES, il ne semble pas y avoir phénomène de conjugaison entre les électrons π du cycle phényl et ceux de la chaîne diènique centrale : la liaison C_1-C_7 de 1,49$_9$ Å correspond à une longueur normale. D'ailleurs l'angle d'un phényl avec le plan moyen de la chaîne centrale est plus élevé que dans le DES (85°9).

La chaîne centrale est à peu près plane et on observe une légère conjugaison entre les électrons π : la liaison C_7-$C_{7'}$ est légèrement plus courte (1,47$_2$ Å) rappelant ce que l'on observe dans le butadiène.

Enfin comme dans le DES, le méthyl terminal C_9 est en contact avec les atomes du cycle phényl adjacent C_2 (3,62$_2$ Å) et surtout C_6 (3,54$_8$ Å).

I-6 CONCLUSION

Contrairement à ce qu'ont pu penser pendant très longtemps les physico-chimistes et les théoriciens, le diéthylstilboestrol peut avoir 2 conformations différentes d'énergie voisine. Il est même très vraisemblable que la forme pseudo binaire apparaisse dès la solution et ne soit pas seulement le résultat des contraintes imposées par le champ cristallin.

Par contre, le cycladiène constitue une molécule rigide; l'introduction d'une double liaison supplémentaire dans la chaîne aliphatique centrale empêche pratiquement toute déformation importante de la molécule.

Après avoir prouvé leur existence, il peut être intéressant de comparer le comportement de chacune des formes de diéthylstilboestrol par rapport à celui des stéroïdes oestrogènes, dans le milieu cristallin. C'est ce que nous allons faire maintenant.

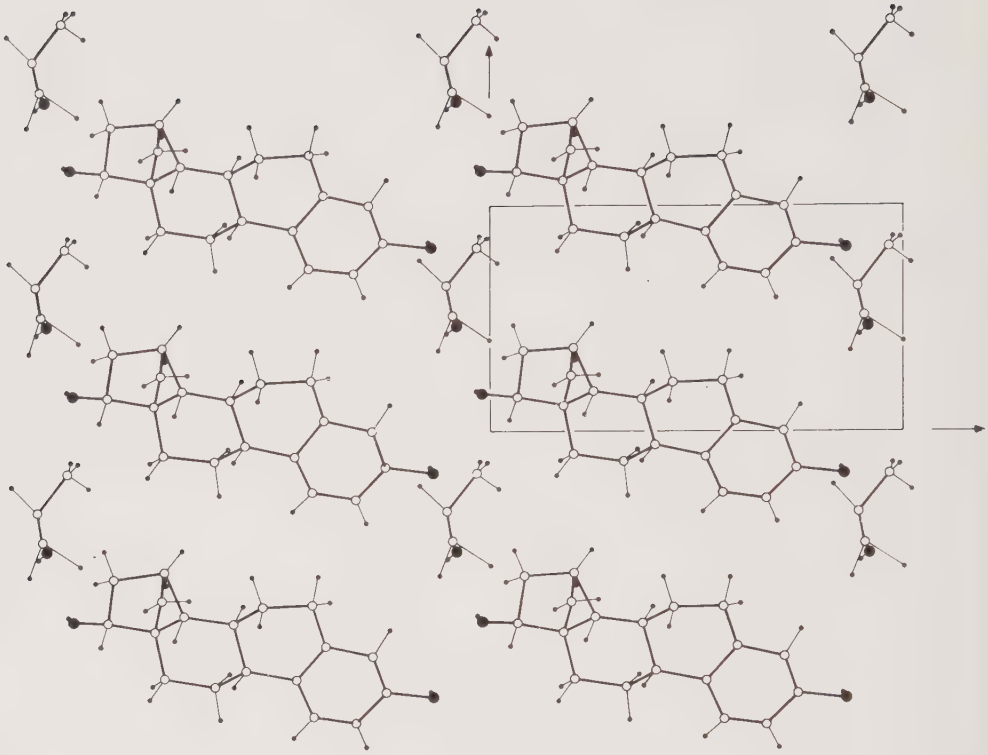
C H A P I T R E II

COMPARAISON DE LA COHESION CRISTALLINE DANS LES STRUCTURES DU DES ET DE L'OESTRADIOL

Pour que deux composés aient même activité, il faut généralement qu'ils présentent une certaine similitude de conformation et de champ moléculaire. Sans toutefois aller jusqu'à reprendre l'idée, émise par de nombreux auteurs (notamment H.E.Ungnade (20)), que de tels composés sont susceptibles de syncristalliser, il nous parait nécessaire, après avoir mis en évidence pour le DES, l'existence de plusieurs conformations, de comparer leur comportement dans le milieu cristallin avec celui de l'oestradiol. On peut penser que celle qui présentera l'assemblage moléculaire le plus proche de l'oestradiol cristallisé, aura de fortes chances d'être la forme biologiquement active.

La cohésion cristalline des composés oestrogènes stéroïdiques a déjà fait l'objet de plusieurs études de notre part (5a , 5b), nous rappellerons seulement ici, les principales caractéristiques de l'assemblage moléculaires dans les différentes structures de l'oestradiol.

Les 2 seules structures connues de l'oestradiol (hémihydratée et complexée avec l'alcool propylique) appartiennent toutes deux au système orthorhombique $P2_12_12_1$ ou $P2_12_12$ (voir les 2 projections xyO décrites en annexe). Les molécules sont organisées en feuillets parallèles au plan (a,c). La cohésion à l'intérieur de ces feuillets est réalisée par l'intermédiaire d'un réseau de liaisons hydrogène parallèlesà la direction c.



Dans l'oestradiol hémihydratée chaque molécule d'eau appartient en fait à deux feuillets (a.c) adjacents et l'on peut dire que le réseau de liaisons hydrogène crée un feuillet bimoléculaire.

Les feuillets ainsi constitués ne sont liés entre eux que par des interactions de Van der Waals de faible énergie.

II-1. LES STRUCTURES ORTHORHOMBIQUES Pbca DU DES ET DU CYCLADIENE

Principal oestrogène de synthèse, le diéthylstilboestrol possède donc au moins deux configurations moléculaires différentes suivant le mode de cristallisation. La forme centrosymétrique se rencontre dans la structure non solvatée (obtenue par sublimation ou à partir d'une solution de DES dans l'hexane); la molécule est située sur un centre de symétrie. Le groupe spatial est orthorhombique Pbca. Le cycladiène pour lequel seule la configuration centrosymétrique peut exister, cristallise dans le même groupe spatial quel que soit le milieu de cristallisation.

Dans les 2 cas, on observe un assemblage en chevrons.

II-1-1 Etude de la cohesion cristalline suivant b,c

On peut considérer que les forces de cohésion du cristal les plus fortes sont situées dans ces 2 directions bien qu'elles ne soient dues en grande partie qu'à des liaisons hydrogène de faible énergie analogues à celles rencontrées pour les stéroïdes androgènes(voir 2ème Partie).Les distances entre les deux atomes d'oxygène intéressés sont en effet très élevées :

3,03 Å dans le diéthylstilboestrol et même 3,06 Å dans le cas du cycladiène ; comme pour les stéroïdes androgènes, il est cependant possible de continuer à parler de liaisons hydrogène d'autant plus que les angles de liaison O-H...O correspondants sont très voisins des valeurs habituellement rencontrées (respectivement 171° et 167°).

Ces liaisons hydrogène forment des chaînes brisées allongées suivant Oz, et constituent une sorte de canevas réunissant les molécules de diéthylstilboestrol en "feuillet gaufré" (b,c). La direction moyenne des molécules n'est pas parallèle au plan du feuillet mais plutôt à la direction (110) par la suite on obtient un feuillet dont l'épaisseur représente la moitié du paramètre a.

II-1-2 Cohésion cristalline suivant a

La cohésion cristalline entre ces feuillets est assurée uniquement par des forces de dispersion de faible énergie qui relient deux molécules de diéthylstilboestrol symétriques par rapport à l'axe hélicoïdal 2_1 parallèle à Oz. Ces interactions se situent essentiellement entre cycles phényl du DES qui ne sont pas disposés parallèlement l'un à l'autre comme l'empilement le plus compact pourrait le suggérer mais plutôt presque perpendiculairement (on observe un angle égal à 70°1dans le DES et à 69°2 dans le cycladiène) comme cela a lieu par exemple pour le naphtalène. Nous retrouverons la même disposition des cycles phényles dans les formes triclniques que nous allons décrire maintenant.

II-2 LES STRUCTURES SOLVATEES DU DES DE SYMETRIE TRICLINIQUE

Le diéthylstilboestrol cristallisé dans un solvant polaire adopte généralement la configuration à symétrie pseudo binaire que nous avons décrite au chapitre précédent. Les structures cristallines que nous avons étudiées jusqu'ici (méthanol-eau, éthanol, DMSO) appartiennent toutes au groupe spatial $P\bar{1}$ (*).

(*) Il existe des formes solvatées correspondant à des groupes de symétrie différents que nous étudierons vraisemblablement dans l'avenir (avec le propanol on obtient une structure monoclinique $P2_1/c$, avec le dioxane le groupe de symétrie est P1).

Elles sont toutes plus ou moins instables : les deux formes cristallines alcoolées isolées de la liqueur mère,à l'air libre,se détruisent très rapidement à la température ambiante (en moins d'une heure) pour donner une poudre microcristalline de symétrie orthorhombique Pbca.Le cristal obtenu dans le diméthyl sulfoxyde est nettement plus stable et peut être conservé au moins une journée.

Dans ces structures,des liaisons hydrogène nettement plus énergétiques que dans le cas précédent étirent les molécules de stilboestrol en feuillets parallèles. On observe donc pour la configuration pseudo binaire un arrangement cristallin qui rappelle la disposition des molécules d'oestradiol dans les structures déjà étudiées.

Bien que la symétrie finale de ces structures soient identiques et les mailles voisines, la cohésion cristalline est réalisée de façon fort différente, c'est pour cette raison que nous étudierons chaque structure séparément.

II-2-1 DES + CH_3OH + H_2O

Dans cette structure, l'ensemble des liaisons hydrogène assure la cohésion dans un plan (a,b). Elles constituent une chaîne brisée parallèle à Ox. Dans la direction c, ce sont uniquement des forces de Van der Waals qui assurent la construction de l'édifice cristallin. Comme dans les structures de l'oestradiol, chaque atome d'oxygène d'une molécule de DES est lié à des atomes d'oxygène de deux autres molécules :

- Une des liaisons s'établit directement entre atomes d'oxygène phénolique :

O_{10} - H ... O_{10} = $2,75_9$ Å ; l'angle de la liaison est 170°

- l'autre liaison entre oxygène phénolique est assurée par l'intermédiaire des molécules de solvates et comprend trois liaisons hydrogène successives :

- l'atome O_{10} est très fortement lié à la molécule de méthanol :

O_{10} - H ... $O_{méthanol}$ = $2,60_6$ Å ; l'angle de la liaison est 148°

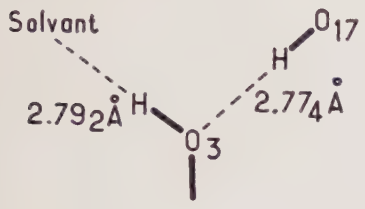
- puis une liaison de même intensité lie entre elles les deux molécules de solvates : $O_{méthanol}$ ... W = 2,63 Å (ces molécules ont une forte agitation thermique anisotrope - principalement suivant Oz - et il ne nous a pas été possible de placer les atomes d'hydrogène engagés dans ces liaisons).

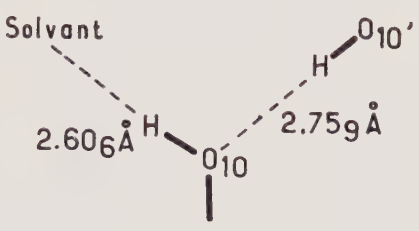
- enfin une liaison hydrogène très faible ($3,00_6$ Å) lie la molécule d'eau à l'atome d'oxygène $O_{10'}$.

Remarquons tout d'abord que la molécule d'eau n'est liée que 2 fois. A partir des 2 atomes d'hydrogène qui ont pu être placés, on peut supposer qu'elle reçoit un hydrogène du méthanol et en engage un dans la liaison avec le diéthylstilboestrol. Il y a donc un atome d'hydrogène de l'eau qui n'est pas lié. Cette très faible fixation de la molécule d'eau explique

la fragilité de l'édifice cristallin (une étude à l'ATD montre que cette molécule est très facilement extraite du cristal, avant la molécule de méthanol plus fortement liée (4)).

Si on compare cette structure avec celle de l'oestradiol hémihydratée (où nous avons pu également localiser les atomes d'hydrogène engagés dans des liaisons) on s'aperçoit que l'atome d'oxygène O_{10} donne des liaisons semblables à cellesde l'atome d'oxygène phénolique O_3 de l'oestradiol :





Dans les deux cas il donne son hydrogène au solvant avec établissement d'une liaison forte, et présente un doublet électronique à l'hydrogène de l'autre fonction-ol de la molécule d'oestrogène.

D'autre part,les liaisons de cette deuxième fonction-ol avec le solvant sont nettement plus faibles dans les deux cas. (3,00$_6$ Å pour le $O_{10'}$ du DES, 2,85$_5$ pour le 17 β ol de l'oestradiol). Finalement,on peut penser que l'oxygène O_{10} a conservé intégralement sa fonction phénolique et s'identifie à l'oxygène O_3 correspondant de la molécule d'oestradiol,par contre la fonction de l'atome $O_{10'}$ est considérablement affaiblie et s'apparente plutôt à la fonction alcool tertiaire O_{17} de l'oestradiol.

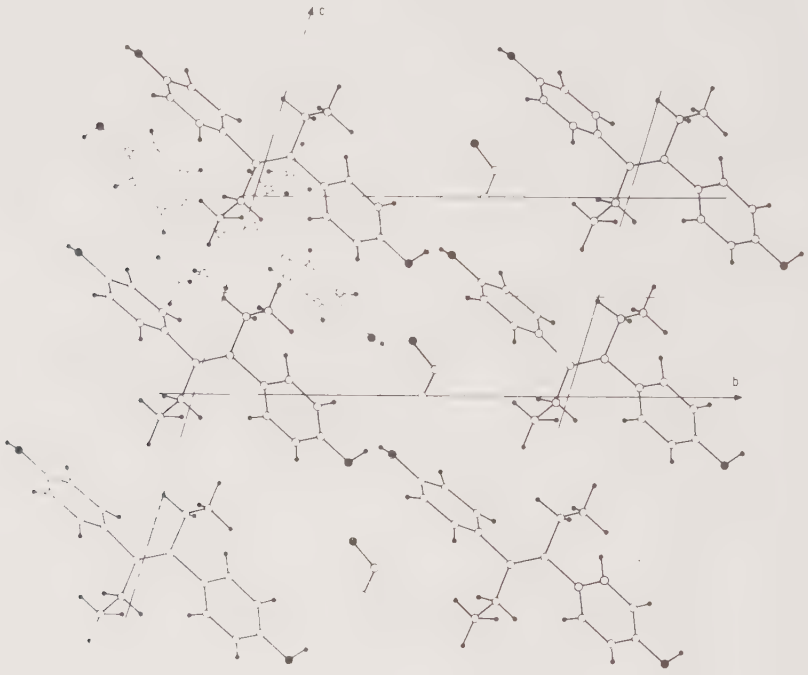
Par analogie avec la molécule stéroïdique, nous appellerons cycle A dans le DES pseudo binaire, le cycle benzénique qui porte l'atome d'oxygène O_{10}.

Dans la direction c la cohésion cristalline est assurée par des contacts de Van der Waals entre molécule de DES symétriques par rapport à un centre de symétrie. On remarquera notamment des contacts entre le méthylène $C_{(8')}H_2$ (voir figure en annexe) et le groupement C_6H du cycle A(distances C-C 3,70 Å) et avec un autre méthylène $C_{(8')}H_2$ (3,75$_8$ Å). Il y a également un contact assez prononcé entre l'atome d'oxygène du méthanol et l'atome de carbone C_5 (3,35 Å).

II-2-2 DES + C_2H_5OH

Entre les deux formes alcoolées du DES, on observe des paramètres cristallins très voisins et l'arrangement général des molécules de stilboestrol est à première vue identique, mais les forces qui assurent la cohé-

sion cristalline sont totalement différentes. Ici les liaisons hydrogènes forment une spirale parallèle à la direction Oz : leur disposition rappelle encore davantage celle que nous avons rencontré pour les structures de l'oestradiol et nous pouvons parler comme nous l'avons fait alors de feuillet (b,c). D'ailleurs,comme pour l'oestradiol hémihydraté, les liaisons hydrogène ont constitué un feuillet bimoléculaire : il regroupe deux couches de molécules parallèles. Nous avons reconstitué une de ces couches sur la figure ci-dessous.



La cohésion du feuillet est assurée par des liaisons fortes

liaisons hydrogènes

-et par desinteractions de Van der Waals (les molécules de DES ayant rigoureusement la même orientation que dans la structure précédente on retrouve ici les mêmes interactions que celles décrites plus haut).

Ici aussi, l'atome d'oxygène O_{10} donne une liaison forte avec l'éthanol ($2,71_8$ Å) alors que l'atome $O_{10'}$ donne une liaison très faible ($2,99_1$ Å). La liaison $O_{10} \ldots O_{10'}$ a une longueur nettement plus importante que dans la structure précédente ($2,84_4$ Å).

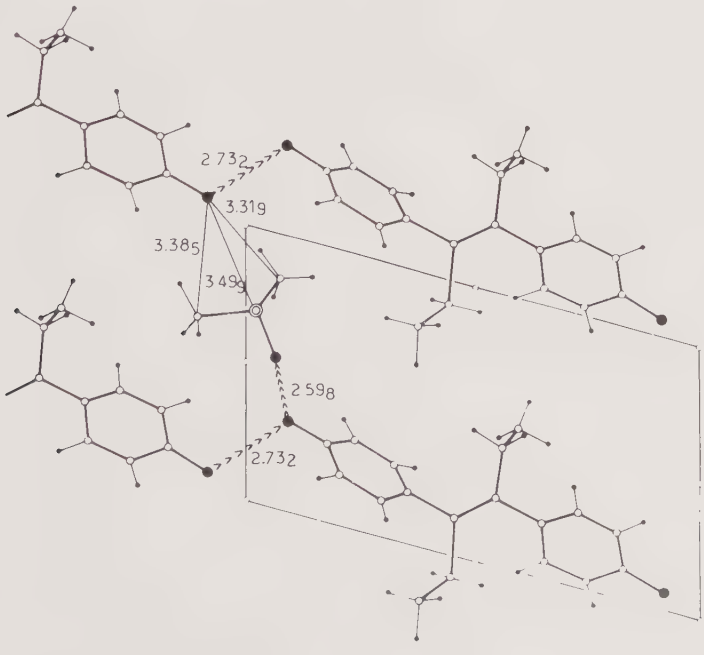
Les différents feuillets bimoléculaires (b,c) ne sont associés entre eux que par des interactions de Van der Waals de faible énergie, principalement au niveau des molécules d'éthanol : l'extrémité de cette molécule est liée à une autre molécule d'éthanol ($3,76_7$ Å) ainsi qu'à l'atome $C_{3'}$ du cycle phénylique ($3,65_2$ Å).

II-2-3 DES + Diméthyl sulfoxyde

Il est assez surprenant de constater que cette structure est beaucoup plus stable que les précédentes, alors qu'il n'y a presque plus de liaisons hydrogène pour assurer la cohésion cristalline.

- l'atome d'oxygène O_{10}, qui donnait jusqu'ici des liaisons hydrogène de faible énergie avec la molécule de solvate, n'engage plus aucun de ses doublets électroniques dans une liaison. Par la suite, si l'on considère l'ensemble des liaisons hydrogène, la différence entre les 2 atomes d'oxygène est encore plus marquée, puisque l'atome O_{10} est au contraire très fortement lié, ici, à l'oxygène de la molécule de DMSO ($2,59_8$ Å). Peut-on rendre cette différence énergétique plus importante entre les 2 atomes d'oxygène, responsable de l'augmentation de l'écart observé entre le plan des phényls du DES dans ce cas (70° au lieu de 60° dans les 2 structures précédentes). Nous disposons de trop peu d'éléments pour pouvoir l'affirmer avec certitude mais il pourrait très bien y avoir une relation entre ces deux phénomènes.

Les liaisons hydrogène reliant entre elles les molécules de DES, forment un enchaînement moléculaire parallèle à la direction b. Par contre, elles n'assurent pas directement la cohésion dans la direction c. Les molécules de DMSO fixées aux enchainements de DES à l'aide d'une forte liaison hydrogène, maintiennent l'existence du feuillet (b,c) décrit précédemment, par des contacts de Van der Waals de forte énergie avec l'atome d'oxygène O_{10} de la chaîne située directement en dessous. Cet assemblage particulier du feuillet (b,c) est représenté sur la figure ci-dessous où l'on remarquera les contacts des méthyls du DMSO avec l'oxygène O_{10} ($3,31_9$ Å et $3,38_5$ Å) et surtout le contact soufre-oxygène à $3,49_9$ Å.



Cette constitution légèrement différente du feuillet (b,c) explique en grande partie les variations des paramètres cristallins par rapport aux structures précédentes et notamment la fermeture de l'angle γ (88°8 au lieu de 98°6 et 100°9).

Dans la direction a, la cohésion est assurée par de multiples interactions de Van der Waals sans qu'il y ait toutefois de contacts remarquables.

II-3 CONCLUSION

Avant de terminer ce chapitre nous voudrions une nouvelle fois souligner quelques traits particuliers de la cohésion cristalline des structures du diéthylstilboestrol.

Quand les 2 atomes d'oxygène du DES sont dans un état électronique identique, c'est à dire quand les molécules environnantes donnent le même type de liaisons hydrogènes, les molécules de DES adoptent une disposition en chevrons ne présentant aucune ressemblance avec l'arrangement cristallin rencontré pour un stéroïde oestrogène.

Au contraire dès l'instant où chacun des atomes d'oxygène est traité de façon différente par le milieu ambiant, la molécule de DES adopte une conformation qui à priori semble moins stable que la forme centrosymétrique, mais qui donne des assemblages cristallins ressemblant à ceux rencontrés pour l'oestradiol avec notamment :

- une disposition en feuillets parallèles où la cohésion est assurée principalement par liaisons hydrogènes.

- et dans la direction c un encombrement moléculaire à peu près voisin (la distance de 2 plans 001 et 002 du réseau triclinique qui est égale à $\frac{V}{|\vec{a} \wedge \vec{b}|} \simeq 7{,}3$ Å peut être comparée aux 6,63 Å du paramètre c dans l'oestradiol hémihydraté).

Cette conformation du DES n'est stable qu'à la condition de maintenir le DES en présence du solvant. Les formes cristallines triclinique obtenues ne nous semble pas stable parce que l'on ne peut maintenir en place la molécule du solvant : à l'air libre on a un système instable constitué du cristal solvaté et du milieu ambiant où la tension de vapeur du solvant est pratiquement nulle; il a donc tendance à quitter le milieu cristallin). Mais rien ne prouve que le récepteur ne possède pas d'éléments qui, jouant le rôle de la molécule de solvant stabiliserait la forme pseudo-binaire.

CHAPITRE III

RELATION STRUCTURE-ACTIVITE DES-OESTRADIOL

III-1 AVANT-PROPOS

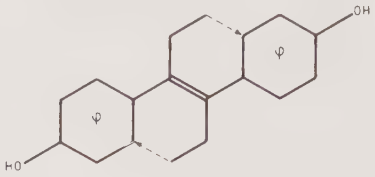
Partant de l'identité d'action et de fixation du DES et de l'oestradiol vis à vis du récepteur utérin, nous nous sommes proposés, en travaillant sur les modèles stéréochimiques fournis par l'analyse radiocristallographique, d'améliorer la théorie des cycles potentiels de Dodds basée uniquement sur une analyse bidimensionnelle. Il semblerait que l'application, avec succès, de cette théorie à la réalisation d'oestrogène de synthèse comme le DES, le cycladiène ne soit que le produit d'un heureux hasard. Nous avons en effet montré que les groupements éthyles ne pouvaient, en aucune façon, être considérés comme les esquisses des cycles B et C du stéroïde. Il est vraisemblable que leur rôle est totalement différent.

Le problème posé par le DES est en réalité plus complexe. Nous avons montré qu'il existe en réalité trois conformations possibles :

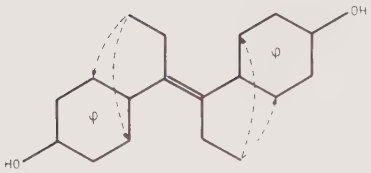
- une forme centrosymétrique $DES_{\bar{1}}$
- deux formes à symétrie pseudobinaire, isomères optiques l'une de l'autre, $DES_{"2"(g)}$ et $DES_{"2"(d)}$.

De plus, s'il existe des contacts relativement étroits entre les carbones terminaux des groupements éthyls et les cycles aromatiques

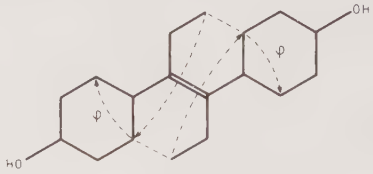
adjacents, on observe en réalité un double contact pour chacun de ces atomes et ces divers contacts ne correspondent absolument pas à l'hypothèse émise par Dodds :



DES vu par Dodds



$DES_{\bar{1}}$



$DES_{"2"}$

Comme il n'est pas pensable que ces 3 formes différentes soient également actives, il serait souhaitable de trouver celle qui constitue la véritable hormone oestrogène. A partir des connaissances biochimiques (*) et cristallographiques, dont nous disposons actuellement, il est peut être possible d'identifier "l'hormone oestrogène diéthylstilboestrol" à l'une des 2 conformations $DES_{\bar{1}}$ ou $DES_{"2"}$.

Avant d'entreprendre cette discussion, il nous semble cependant nécessaire de rappeler, encore une fois, que le problème de la stabilité relative des 2 conformations du diéthylstilboestrol tel qu'il apparaît à travers les structures cristallines étudiées, est en réalité un faux problème : si les formes $DES_{"2"}$ sont peu stables, c'est à cause de la fugacité des molécules de solvant ; le solvate détruit, les molécules reprennent la conformation centrosymétrique. Le calcul de potentiel moléculaire, présenté au premier chapitre, montre au contraire que la forme $DES_{"2"}$ est vraisemblablement aussi stable; on peut d'ailleurs imaginer qu'il existe, au niveau de l'organisme des résidus moléculaires capables de jouer le rôle de la molécule de solvant, et donc d'induire la conformation $DES_{"2"}$ du diéthylstilboestrol. Vu la faible différence d'énergie existant entre les 2 conformations, il n'existe, à priori, aucune raison d'affirmer que, dans un organisme vivant, une forme sera plus stable que l'autre.

III-2 ACTIVITE POSSIBLE DU $DES_{\bar{1}}$

Considérons tout d'abord la forme centrosymétrique, qui semble se rapprocher davantage de la formule classique.

(*) Les valeurs des différentes constantes d'association, utilisées dans ce chapitre, nous ont été fournis par le Laboratoire des Hormones du Professeur BAULIEU (INSERM - Bicêtre).

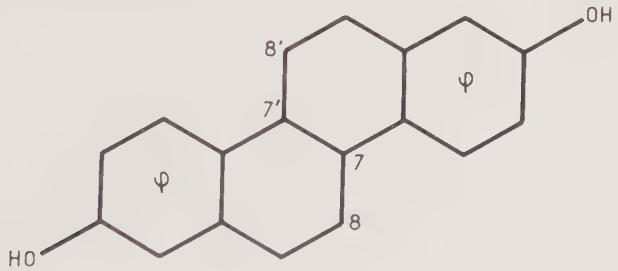
Pour ce composé, la cristallographie donne raison à l'hypothèse de HP.Koch. Les cycles aromatiques sont parallèles et forment un angle de 62°8 avec la chaîne aliphatique centrale, mais le fait que cette chaîne soit effectivement dans une configuration trans-trans constitue, à notre avis, le premier renseignement important : son extension est alors maximale. Par la suite, l'encombrement de la molécule de $DES_{\bar{1}}$ (environ 5,8 Å dans la direction perpendiculaire aux plans des phényls) est nettement supérieure à celui d'un oestrogène stéroïdique (3,4 Å dans la même direction).

Un autre détail surprend dans cette conformation, c'est la faiblesse des liaisons hydrogène fournies par les atomes d'oxygène terminaux. Nous pouvons nous demander si ces liaisons de faible énergie sont la conséquence de l'action du champ cristallin, ou si elles constituent, au contraire, une propriété caractéristique de cette conformation. Tant que nous ne possèderons pas une structure de la forme centrosymétrique où la cohésion cristalline sera différente, nous ne pourrons pas répondre avec certitude.

Cependant, la symétrie des deux extrémités phénoliques constituent vraisemblablement le point le plus important. Vue sous cet aspect, cette conformation s'éloigne assez nettement du stéroïde oestrogène, type "oestradiol," dont :

- l'une des fonctions hydroxyles (O_3) est effectivement un phénol
- mais où l'autre fonction (O_{17}) est un alcool tertiaire

La conformation centrosymétrique se rapprocherait par contre davantage de la molécule suivante



qui, comme le stéroïde, constitue un ensemble rigide. C'est un composé admettant un centre de symétrie, les cycles aromatiques sont parallèles, et on peut supposer que les atomes d'oxygène terminaux ont une structure électronique différente comme dans le $DES_{\bar{1}}$.

Nous avons construit cette molécule en la considérant comme formée de 2 parties symétriques constituées chacune par les cycles A et B observés dans l'oestradiol. La distance entre les 2 atomes d'oxygène est égal à 12,06 Å alors qu'elle est de 12,13 Å dans le DES. Les cycles aromatiques font un angle de 51° avec le plan de la "chaîne centrale", c'est à dire avec le plan moyen des atomes C_7-$C_{7'}$, C_8 et $C_{8'}$,(cf. notation DES).

Or la constante d'association de ce composé avec le récepteur utérin est environ 100 fois plus faible que pour l'oestradiol, alors que le diéthylstilboestrol se lierait mieux que ce dernier avec le récepteur utérin. (Il faut cependant atténuer légèrement cet argument défavorable à la forme $DES_{\bar{1}}$, en remarquant que nous avons raisonné ici, comme si le DES se cyclisait dans cette conformation, comme l'avait prévu Dodds, or nous avons déjà signalé qu'il n'en était pas ainsi.

Enfin nous avons vu au chapitre I que la présence de 2 doubles liaisonsconjuguées sur la chaîne centrale, rendait la molécule de cycladiène pratiquement indéformable. Comme le DES_1 elle possède une conformation centrosymétrique; or sa constante d'affinité pour le récepteur utérin est au moins 10 fois plus faible que celle du DES ou de l'oestradiol.

Pour toutes ces raisons, nous ne pensons pas que le DES_1 puisse représenter la forme active du diéthylstilboestrol. Nous allons voir maintenant qu'au contraire la conformation pseudo-binaire pourrait fort bien constituer l'hormone oestrogène.

III-3 ACTIVITE POSSIBLE DU DES"2"

Nous noterons tout d'abord que l'encombrement du DES"2" dans la direction perpendiculaire à l'un des cycles aromatiques, maintenant égal à 4,8 Å, reste toujours supérieur à celui d'un oestrogène stéroïdique. Il est cependant nettement plus petit que celui de la molécule centrosymétrique.

De plus, l'état électronique des atomes d'oxygène phénoliques est certainement différent, comme le montre leur aptitude différente à former des liaisons hydrogène. L'atome d'oxygène O_{10} donne des liaisons hydrogène d'énergie comparable à celle des liaisons obtenues pour l'atome d'oxygène O_3 dans les différentes structures d'oestradiol étudiées.

Nous pouvons par la suite considérer qu'il conserve son caractère phénolique, et, par analogie avec la nomenclature utilisée pour les stéroïdes, nous appellerons cycle A, le cycle benzénique qui le porte.

Tableau comparatif des liaisons hydrogène rencontrées dans les structures de l'oestradiol et du DES "2"

(S désigne suivant le cas une molécule d'eau ou un hydroxile alcoolique)

	Oestradiol hémihydraté	Oestradiol +alc propyl.	DES+CH_3OH +H_2O	DES +C_2H_5OH
$O_{3(10)}\cdots O_{17(10')}$	$2,77_4$ Å	$2,75_9$ Å	$2,75_9$ Å	$2,84_4$ Å
$O_{3(10)}\cdots S$	$2,79_2$ Å	$2,63_6$ Å	$2,60_6$ Å	$2,71_8$ Å
$O_{17(10')}\cdots S$	$2,85_5$ Å	$2,72_4$ Å	$3,00_6$ Å	$2,99_1$ Å

Comme dans l'oestradiol, où l'oxygène O_{17} donne des liaisons hydrogène comparativement plus faibles que l'oxygène phénolique, l'atome

d'oxygène O_{10}, donne également des liaisons beaucoup plus faibles que l'oxygène phénolique O_{10}. On pourrait admettre qu'il a perdu, en partie, son caractère phénolique, et qu'il se rapproche davantage de la fonction alcool tertiaire rencontrée dans l'oestradiol. Les molécules de DES$_{"2"}$ constituent donc, au point de vue structure électronique des groupements oxygénés, une meilleure approche de la molécule d'oestradiol, que ne le fait la forme centrosymétrique que nous avons vu précédemment.

A cette meilleure ressemblance au point de vue de l'encombrement moyen et de la structure électronique des atomes d'oxygène, correspond en plus une certaine ressemblance de l'arrangement cristallin : nous avons pu voir que, comme dans les différentes structures de stéroïdes oestrogènes ($P2_12_12_1$ ou $P2_1$), les molécules de DES$_{"2"}$ adoptent une disposition en feuillets parallèles à l'intérieur desquelles la cohésion est assurée principalement par des liaisons hydrogène (On pourra comparer, par exemple, les projections des structures de l'oestradiol hémihydraté et du DES + éthanol, représentées en annexe).

Les paramètres cristallins définissant les feuillets dans ces deux cas ont également des dimensions comparables :

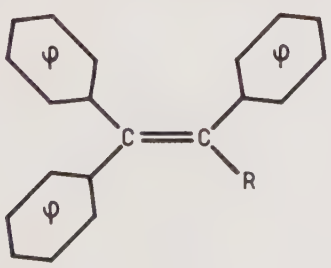
$a = 12{,}055$ et $12{,}215$ Å ; $c = 6{,}63_2$ et $6{,}67_1$ Å pour l'oestradiol

$b = 13{,}657$ et $13{,}84_1$ Å ; $d = \dfrac{V}{|\vec{a} \wedge \vec{b}|} \simeq 7{,}3$ Å pour le DES

Du point de vue physico-chimique la conformation DES$_{"2"}$ est donc beaucoup plus proche de l'oestradiol que la forme centrosymétrique. Il semble que l'on disposerait d'un argument décisif, si l'on pouvait trouver une molécule rigide, de conformation très voisine de celle du DES$_{"2"}$ et possédant une constante d'affinité élevée pour le récepteur utérin. Les dérivés du triphényléthylène, dont l'activité a longtemps intrigué les biochimistes, peuvent fournir cet argument.

III-4 LE PROBLEME DES DERIVES DU TRIPHENYLETHYLENE

Ce problème a été soulevé pour la première fois par J.Jacques dans ses "mises au point" (11) , l'activité oestrogène des dérivés du triphényléthylène lui semblait en effet incompréhensible à la lumière de la



théorie des cycles potentiels de Dodds. Dans ces composés on ne retrouve pas d'éléments pouvant esquisser les cycles B et C du stéroïde. Ce phénomène ne nous surprend plus tellement maintenant que nous avons démontré que la notion de cycles potentiels ne peut pas s'appliquer dans le cas du diéthylstilboestrol.

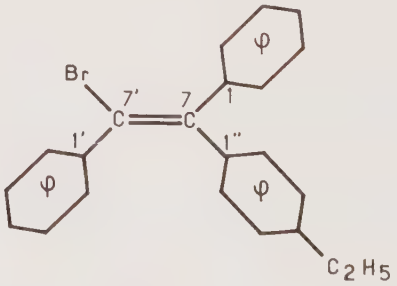
Cependant les dérivés du triphényléthylène nous posent d'autres problèmes intéressants. Il est par exemple assez surprenant de voir que le triphényléthylène (R = H) est complètement inactif, mais que les dérivés où le substituant R est assez volumineux, constituent au contraire des oestrogènes très actifs (R = CH_3, C_2H_5 ou Br).

Considèrant uniquement le dérivé bromé, J.Jacques a estimé que l'électronégativité de cet atome pouvait induire une hydroxylation in vivo en para sur les cycles aromatiques. En effet, ces dérivés ne sont généralement pas de véritables oestrogènes mais agissent plutôt comme des pro-oestrogènes (substances susceptibles de se transformer dans l'organisme en oestrogène).

Cependant à notre avis cela ne recouvre pas l'ensemble du problème et le fait que les dérivés méthylés et éthylés puissent fournir après hydroxylation des substances également actives alors que le dérivé chloré n'est pas actif, permet de proposer une hypothèse légèrement différente.

Ces groupements différents au point de vue électro négativité, présentent un encombrement à peu près identique : ils empêchent donc la libre rotation du cycle phényl adjacent. Par la suite il n'existe pas sans doute beaucoup de conformations différentes pour l'ensemble des 3 cycles et il est très vraisemblable que ces dérivés bromés ou méthylés constituent des molécules relativement rigides. Elles présentent le même intérêt que le cycladiène; il serait donc intéressant de savoir quelle est la conformation des cycles en position trans.

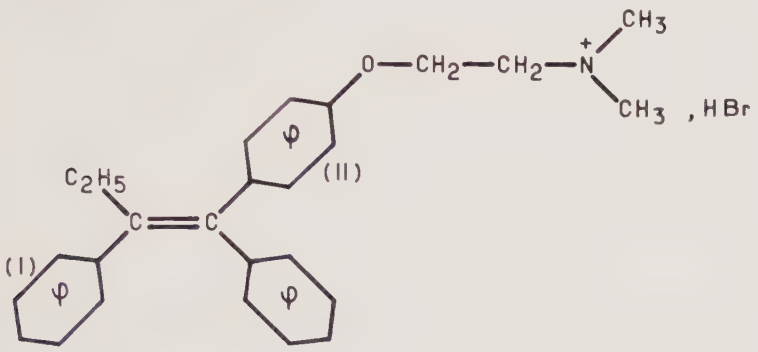
Dans ce but nous avons étudié la structure du dérivé bromé suivant :



qui est également un pro-oestrogène.

Dans cette structure les cycles phényls en positions trans sont croisés comme dans la molécule de DES"2"; ils font avec le plan moyen des atomes centraux (C_7-$C_{7'}$ - C_1-$C_{1'}$ - $C_{1''}$ et Br) des angles respectivement égaux à 65°3 et 53°4.

Un résultat identique avait déjà été obtenu par B.T.Kilbourn et P.G.Owston (12) pour le dérivé :



Ce composé (I.C.I. 47499) est au contraire un oestrogène véritable; il possède toutefois une constante d'association avec le récepteur utérin légèrement plus faible que celles de l'oestradiol et du DES.

Il cristallise avec 2 molécules dans l'unité asymétrique et pour chacune de ces molécules, les cycles phényles adoptent la même configuration que dans le DES"2".

Les angles qu'ils forment avec la chaîne éthylénique centrale, sont du même ordre de grandeur que ceux rencontrés dans le DES.

- 55° et 53° pour le cycle I
- 52° et 60° pour le cycle II

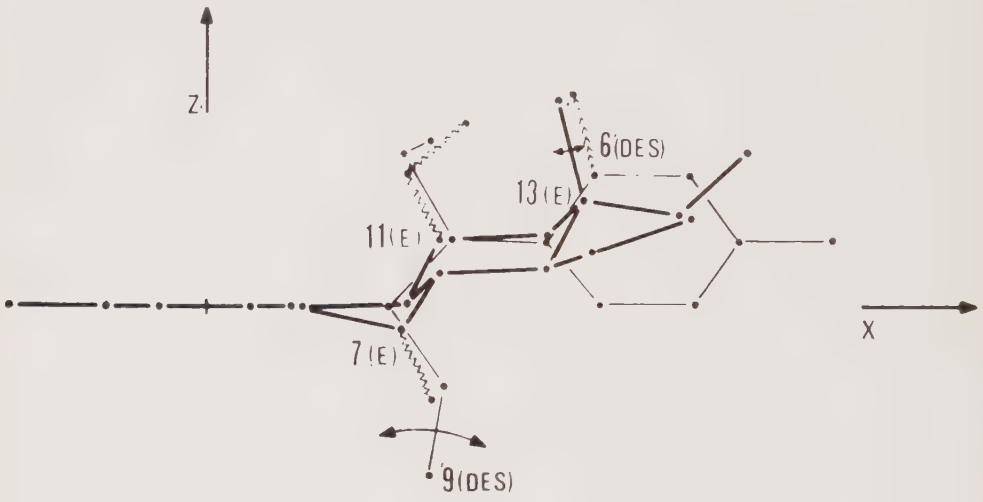
Or, si dans le diéthylstilboestrol il existe encore certains degrés de liberté qui permettent, en solution, le passage de la forme DES-1 à la forme DES"2" les empêchements stériques supplémentaires introduits en remplaçant l'un des éthyls par un cycle phényl, rendent les dérivés du triphényléthylène quasiment indéformables et on peut estimer que la conformation de ces molécules, en solution, est voisine de celles qu'on observe dans le milieu cristallin.

Pour la conformation centrosymétrique du DES, nous avions réuni plutôt des arguments défavorables, ces dernières observations rendent par contre acceptable l'hypothèse que la forme pseudo-binaire représente la forme active du diéthylstilboestrol.

III-5 ESSAIS "D'AMELIORATION" DES MOLECULES D'OESTRADIOL ET DE DES

Considérons maintenant que la molécule de DES"2" représente réellement la molécule active, et comparons là à la molécule d'oestradiol qui

est l'hormone naturelle. Sur la figure ci-dessous nous avons superposé ces deux molécules en prenant comme référence le cycle phénylique A commun. On peut voir que les 2 molécules ne sont pas exactement identiques, ils existent entre elles de nombreuses différences.



Certaines ne pourront être corrigées, c'est par exemple le cas de la différence des écarts entre les groupements oxygénés O_3-O_{17} : dans le diéthylstilboestrol (12,12 Å);il surpasse de plus d'un angstroem celui de l'oestradiol (10,92 Å).

Par contre on remarquera que si les 2 molécules comportent le même nombre d'atomes de carbone, le rôle joué par certains d'entre eux diffère de l'une à l'autre. On peut dès lors espérer obtenir une molécule plus active en complétant chacune d'elles par les groupements qui n'appartiennent qu'à l'autre.

Ainsi si l'on compare la molécule d'oestradiol à celle de diéthylstilboestrol, on voit que l'oestradiol ne possède pas les groupements éthyles du DES ; il serait donc intéressant de les lui donner. On est ainsi amené à considérer les 2 dérivés stéroïdiques suivants :

- le 11 β C_2H_5 oestradiol

et

- le 7 α CH_3 oestradiol

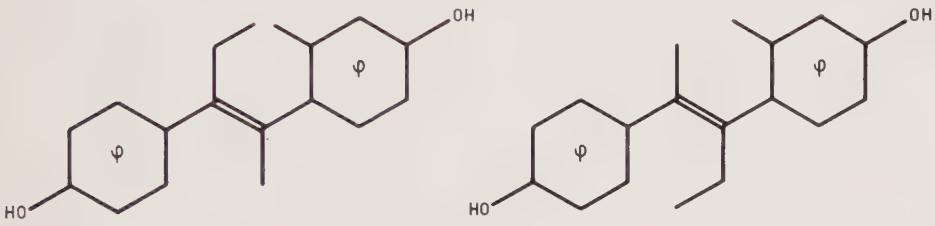
Le dérivé 7 α méthylé (22) s'avère être un composé très actif et sa constante d'association avec le récepteur utérin est égale voire même supérieure à celle de l'oestradiol.

Le dérivé 11 β éthylé n'a jamais été synthétisé, mais, par

contre, on connait le dérivé méthoxylé en 11 β (29) qui doit avoir à peu près la même configuration. Or, alors que le 11 β OH oestradiol n'est pratiquement pas lié au récepteur utérin (ce qui tendrait à prouver que l'oxygène a ici un rôle plutôt néfaste) le 11 β O CH_3 oestradiol est au contraire un oestrogène excessivement puissant.

Ces corrections apportées à l'oestradiol agissant plutôt dans le sens de l'amélioration, nous avons voulu voir si en corrigeant le DES par rapport à l'oestradiol il n'était pas également possible d'obtenir une molécule plus active.

Pour cela il faut réduire les groupements éthyles en 7 et 7' du DES (sans les supprimer complètement de façon à maintenir à peu près la même orientation des cycles benzéniques) et mimer le groupement méthyl C_{18} de l'oestradiol, par un groupement identique en ortho$_{(6')}$ sur le cycle "D" du DES. On obtient alors les deux composés suivants [7] :



Ils ont une constante d'association avec le récepteur utérin au moins égale à celle du DES, alors que les dérivés dépourvus de méthyl nucléaire s'avère nettement moins actif que le DES (au moins 3 fois moins d'après Dodds (7)). L'adjonction en ortho d'un méthyl nucléaire sur le DES semble donc également augmenter l'activité de la molécule.

CONCLUSION

La comparaison des structures et des activités respectives d'un certain nombre de composés oestrogènes rigides nous a permis de supposer que l'hormone diéthylstilboestrol correspondait à la conformation pseudo-binaire. Les dérivés de l'oestradiol où des groupements supplémentaires miment les groupements éthyles du DES s'avèrent également des oestrogènes actifs.

Inversement en essayant de modifier la molécule de DES pour mimer le mieux possible celle d'oestradiol, nous avons pu décrire des composés qui présentent également des constantes d'affinité pour le récepteur utérin, très élevées.

D E U X I E M E P A R T I E

QUELQUES REFLEXIONS SUR LES INTERACTIONS MOLECULAIRES DES COMPOSES OESTROGENES ET ANDROGENES

Dans l'espoir que les interactions moléculaires, constatées au niveau des structures cristallines des molécules de ligand pourrait rendre compte du mode d'interactions avec les protéïnes réceptrices, nous avons, parallèlement à l'étude comparée des molécules d'oestradiol et de DES, déterminé les structures cristallines de différents composés stéroïdes à action oestrogène ou androgène. Le rôle du squelette hydrocarboné ne peut être entrevu dans une telle étude qu'à la condition de disposer de molécules de structures stéréochimiques différentes (comme c'est le cas pour les molécules d'oestradiol et de DES), et il est pratiquement impossible de rechercher, de cette façon, pour qu'elles raisons le squelette stéroïdique est indispensable à l'action androgène.

De ce fait, les quelques particularités que nous voudrions souligner dans les pages suivantes, ont trait essentiellement aux extrémités hydrophiles des hormones stéroïdes.

C H A P I T R E I

ETUDES DES INTERACTIONS MOLECULAIRES AU NIVEAU DE L'OXYGENE O_3 DU CYCLE A

La différence essentielle qui existe entre les stéroïdes oestrogènes et androgènes, se situe au niveau du cycle A, aromatique pour les premiers, cyclohexanique ou hexènique pour les autres.

A priori, on pourrait penser que les différences d'affinités pour les protéines réceptrices sont liées à l'existence d'un phénomène d'interaction entre électrons π,possible dans le cas des oestrogènes, mais qui ne peut se rencontrer au niveau des androgènes. L'étude cristallographique que nous avons entreprise, offre une autre interprétation pour cette différence d'affinité. Elle montre en effet que l'oxygène O_3 du stéroïde fournit des liaisons hydrogène d'intensité différente quand il est porté par un noyau aromatique ou cyclohexanique.

I-1 LIAISONS HYDROGENE FAIBLES (STEROIDES ANDROGENES)

Jusqu'à présent, dans les différentes structures de composés stéroïdes androgènes de la série normale, les liaisons hydrogène intéressant l'atome d'oxygène O_3 sont toujours faiblement énergétiques, aussi bien dans le cas de dérivés 3ol que 3 cétone.

Dans les structures des 2 androstérones, on observe des liaisons où la distance entre atomes d'oxygène est de l'ordre de 3 Å :

$$O_3 - H \overset{164°}{\ldots} O \quad \overset{3,04\ Å}{}$$

pour l'androstérone (dérivé 3α ol) (High 1966)

$$O_3 - H \overset{168°}{\ldots} O \quad \overset{2,95\ Å}{}$$

pour l'épiandrostérone (dérivé 3β ol) (CM Weeks 1971)

Bien que ces 2 composés cristallisent dans le même groupe de symétrie $P2_1$, l'arrangement moléculaire des 2 structures est totalement différent; les liaisons hydrogène observées ne dépendent sans doute pas uniquement de la nature du champ cristallin (*) .

L'influence du champ cristallin sur l'allongement de la liaison hydrogène est encore moins évidente dans les structures des dérivés 3-one que nous avons étudiés : Le carboxyle O_3 est lié à une molécule d'eau W (qui donne également des liaisons hydrogènes avec la fonction 17β ol); pourtant la longueur de la liaison hydrogène, observée dans ce cas, est aussi excessivement allongée :

$$= O_3 \overset{3,02\ Å}{\ldots} H - W$$

pour l'androstanolone

$$= O_3 \overset{2,98\ Å}{\ldots} H - W$$

pour la testostérone

Mais les atomes d'hydrogène n'ont pu être localisés.

(*) *Des liaisons hydroqène faiblement énergétiques ont depuis été également observées par G.Précigoux (24) pour les structures des 3α et 3β 17β androstanediol.*

Malgré la longueur des liaisons observées, il est néanmoins possible de continuer à parler de liaisons hydrogène :

- la limite maximale indiquée par Hamilton (25) pour la distance O...H (3,40 Å) reste encore nettement supérieure à ce que nous observons.

- par ailleurs les angles O-H...O correspondent aux valeurs généralement admises pour une liaison hydrogène (166° ± 9° (25)).

Par contre, dans la structure du 5β androstan 3α 17β diol qui présente une jonction A/B cis, CM Weeks (21_b) observe des liaisons hydrogène plus énergétiques ($2,75_7$ Å et $2,76_3$ Å). Il n'est donc pas possible d'affirmer avec certitude que ces liaisons de faible énergie soient dues à un état électronique particulier de l'atome d'oxygène O_3 plutôt qu'à un simple empêchement stérique.

I-2 LIAISONS HYDROGENE FORTES (STEROIDES OESTROGENES)

Au contraire, les liaisons hydrogènes observées jusqu'ici au niveau de l'oxygène phénolique des composés à action oestrogène semblent nettement plus énergétiques.

Dans les structures de l'oestrone, on a affaire à des liaisons O-H...O de même nature que dans les 2 androstérones (liaison entre un oxygène hydroxylique et un oxygène cétonique), or les longueurs de liaisons observées sont comprises entre 2,78 et 2,83 Å. (La différence entre composés oestrogènes et androgènes apparaît nettement quand on compare les structures de l'oestrone monoclinique $P2_1$ et de l'épiandrostérone; l'analogie certaine qui existe entre les deux arrangements cristallins (molécules disposées en feuillets parallèles) semble exclure l'influence de la stéricité sur les différences de longueurs de liaisons observées).

Dans les 2 structures de l'oestradiol, chaque atome d'oxygène phénolique entre dans deux liaisons hydrogène (une fois comme donneur d'électrons, une fois comme accepteur). Si les liaisons entre 3OH et 17OH sont du même ordre de grandeur que celles observées par Weeks pour le rétroandrostane de ol ($2,77_4$ Å et $2,75_9$ Å), la liaison entre le 3OH et l'hydroxyle de l'alcool propylique (dans la structure oestradiol + propanol) est nettement plus énergétique ($2,63_6$ Å).

I-3 ETUDE DE L'INTERACTION DES - ℓ.TYROSINE ETHYL ESTER

A ce stade de notre étude, nous avons voulu voir si l'existence d'interactions entre électrons π pouvait être responsable de la différence d'affinité des stéroïdes oestrogènes et androgènes pour certaines protéïnes réceptrices. Ce type d'interactions ne peut se produire qu'en présence de certains acides aminés : phényl alanine, tryptophane, tyrosine. La présence de ce dernier acide aminé au niveau du site actif de la Δ_{5-3} oxostéroïde isomérase (dont l'action est inhibée par l'oestradiol et le DES) nous l'a fait choisir pour notre étude.

Nous avons en réalité choisi un dérivé où la fonction acide carboxylique de la tyrosine était estérifiée comme elle l'est dans une protéïne : la ℓ.tyrosine éthyl ester.

Il aurait été souhaitable d'obtenir un cristal du complexe oestradiol + ℓ.tyrosine éthyl ester mais cela n'a pu être possible jusqu'ici.

Par contre le DES semble présenter une grande affinité pour ce résidu protéïque (*) et la cristallisation du complexe n'a pas présenté de difficultés particulières.

Dans cette structure cristalline, le DES adopte une conformation voisine de la forme centrosymétrique décrite précédemment : les deux cycles phényls sont presque parallèles (écart +2°),par contre on note une légère rotation des groupements éthyles (environ 30°). Les deux extrémités phénoliques du DES se comportent presque de façon semblable vis à vis du solvate, bien que les énergies des liaisons respectives soient nettement différentes (**). Dans chacun des cas l'oxygène du DES est lié :

- d'une part au groupement phénolique de la ℓ.tyrosine éthyl ester.
- d'autre part à une fonction amine primaire d'une autre molécule de l'acide aminé.

L'association la plus énergétique est assurée par 2 liaisons hydrogènes relativement courtes :

$2{,}62_5$ Å

Tyr - O-H ..... O-DES

$1{,}30_6$ $1{,}44_2$

$2{,}63_5$ Å

- N - H ..... O-DES

$1{,}28_1$ $1{,}40_9$

Il a été impossible de localiser l'atome d'hydrogène de la fonction phénol du DES, qui semble n'entrer dans aucune liaison. Cependant les autres hydrogènes engagés dans les liaisons étant fortement délocalisés, on peut penser qu'il existe un phénomène de résonance entre les 3 sites d'hydrogène intéressés par cette association.

L'angle du cycle aromatique du DES et du phényl de la tyrosine est égal à 60°8.

Au niveau du deuxième oxygène du DES on retrouve une liaison hydrogène avec le phénol de la tyrosine, mais nettement moins énergétique :

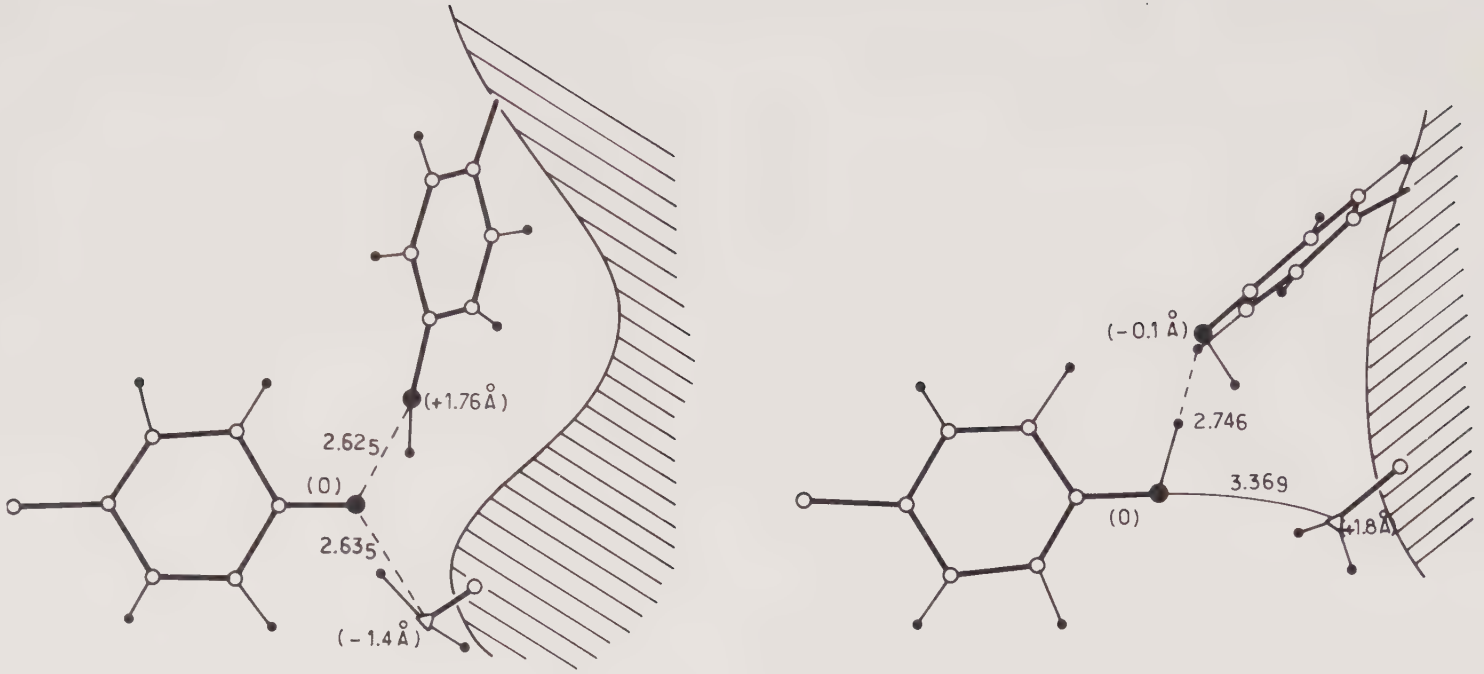
$2{,}74_6$ Å

DES-O-H ..... O-Tyr

$1{,}23_6$ $1{,}53_4$

(*) Une solution d'un mélange en proportion à peu près égales de DES et de ℓ.tyrosine éthyl ester prend très rapidement une légère coloration, preuve indéniable de l'établissement d'interactions fortes entre ces deux molécules.

(**) Cette structure ne peut pas cependant constituer un argument en faveur d'une plus grande activité de la forme du DES à cycles parallèles, les conditions qui régissent la cristallisation du complexe ayant bien peu de chances d'être réalisées au niveau des protéïnes réceptrices.



Orientation des 2 cycles du DES par rapport à la molécule de ℓ tyrosine éthyl ester.

Avec le groupement aminé, il existe seulement un contact de $3,36_9$ Å.

L'angle du cycle aromatique du DES et du phényl de la tyrosine est dans ce cas égal à 88°8.

Cette structure n'apporte peut être pas énormément de renseignements sur la fixation des hormones oestrogènes sur leurs protéïnes réceptrices, car il n'est pas certain du tout qu'un résidu de tyrosine joue un rôle dans l'interaction protéïne-ligand. Cependant, elle rend peu probable l'existence d'interactions entre électrons π (type liaison de "stacking").

Il peut également être intéressant de noter que le groupement $-COOCH_3$ de la ℓ.tyrosine éthyl ester ne joue absolument aucun rôle dans l'association avec le DES.

I-4 QUELQUES OBJECTIONS AUX HYPOTHESES PRECEDENTES

Si, d'après ce que nous avons dit, on admet qu'au niveau de l'oxygène O_3 la différence d'action des oestrogènes et androgènes correspond à une plus ou moins grande aptitude à former des liaisons hydrogènes plutôt qu'à l'existence d'interactions entre électrons π pour les seuls oestrogènes, ceux-ci devraient constituer des inhibiteurs parfaits des récepteurs androgènes; comme ils ne se fixent absolument pas sur ces récepteurs il faut bien reconnaître qu'il intervient d'autres éléments dans le problème de la fixation, et notamment que le rôle du squelette hydrocarboné est loin d'être négligeable.

Par ailleurs, les liaisons hydrogène étant très faibles au niveau de l'oxygène O_3 d'un androgène, on pourrait penser que le stéroïde hormonal se fixe sur les protéïnes réceptrices par la "face 17β ol" de préférence mais la notion de "DES_{n2}" INDUIT" qui a fait l'objet de la IIème partie fait intervenir plus ou moins implicitement une fixation en premier lieu par l'oxygène phénolique.

CHAPITRE II

ETUDE DES INTERACTIONS MOLECULAIRES AU NIVEAU DE LA FONCTION 17β ol

Cette fonction semble jouer un rôle très important, mais encore mal défini, au niveau de différents récepteurs protéïniques de l'organisme vivant :

- elle est absolument indispensable à la fixation sur la protéïne de transport S.B.P. (Steroid Binding Protein).
- le récepteur prostatique présente également une haute spécificité pour les composés possédant une fonction 17β ol.
- enfin si le récepteur oestrogène du cytosol utérin se montre moins intransigeant, il faut reconnaître que le remplacement de cette fonction par un groupement chimique voisin (17one ou 17α ol) diminue considérablement la valeur de la constante d'association.

II-1 STABILITE DE L'ASSOCIATION STEROÏDE-EAU AU NIVEAU DE LA FONCTION 17β ol

Parmi les nombreuses structures cristallines que nous avons étudiées celles des composés possédant une fonction 17 β ol (oestradiol, androstanolone, testostérone) sont les seuls où une molécule d'eau reste emprisonnée. Cette association entre la molécule de stéroïde et l'eau semble très stable ,et l'obtention de formes cristallines anhydres pour ces composés nécessitent des opérations particulières.

- L'oestradiol fourni par le chimiste est toujours une poudre micro cristalline hydratée. Après dissolution dans un solvant rigoureusement anhydre (méthanol, propanol) on obtient la structure hémihydratée que nous avons étudiée. Il faut donc admettre qu'en solution l'eau accompagne la molécule stéroïdique à laquelle elle est liée dans le cristal.

Pour obtenir une forme cristalline déshydratée, il est nécessaire d'opérer sur le produit désséché, sous atmosphère anhydre.

- Dans le cas de l'androstanolone et de la testostérone, les formes hydratéessemblent légèrement moins stables : elles peuvent être déshydratées par chauffage. Remarquons néanmoins que nous les avons cristallisées à partir d'une solution benzénique (!). Pour obtenir une forme anhydre, il faut faire refroidir lentement une solution hexanique de ces composés, réalisée au voisinage de l'ébullition.

II-2 SITUATION DE LA MOLECULE D'EAU PAR RAPPORT AU STEROIDE

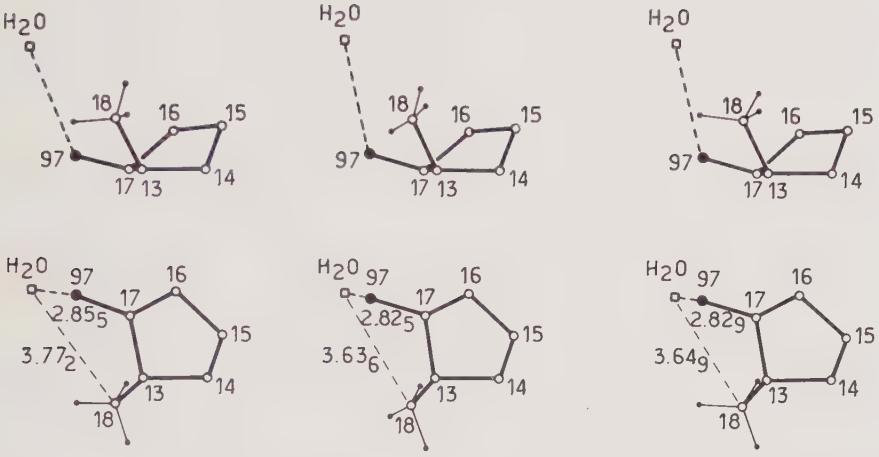
Bien que l'on soit en présence de formes cristallines différentes (orthorhombique $P2_12_12$ pour l'oestradiol; monoclinique $P2_1$ pour les deux androgènes) la localisation de la molécule d'eau par rapport au stéroïde est identique dans les 3 structures étudiées.

La figure ci-dessous représente deux projections orthogonales du cycle D en prenant les atomes C_{13}, C_{14}, et C_{17} pour définir le référentiel.

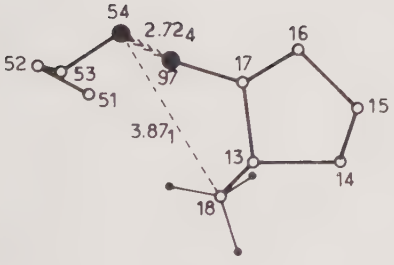
La molécule d'eau est maintenue dans une position bien déterminée.

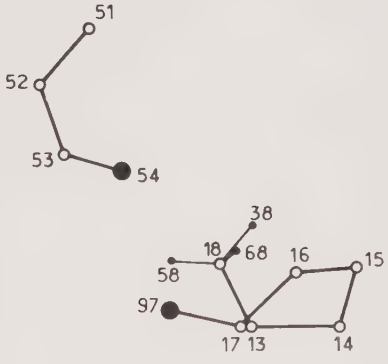
- d'une part par une liaison hydrogène relativement faible avec la fonction 17β ol (elle varie de $2{,}82_5$ à $2{,}85_5$ Å).

- d'autre part par une liaison de Van der Waals faisant intervenir le groupement méthyl C_{18}. (La molécule d'eau vient se loger dans le "sillon" des hydrogènes du groupement méthyl. Les liaisons de valence C-H du carbone C_{18} étant toujours situées dans les plans bissecteurs des angles dièdres définis par la liaison C_{13}-C_{18} et les atomes C_{14}, C_{12} et C_{17}, on comprend mieux que la localisation de l'eau par rapport à la molécule stéroïdique reste constante).



Cette localisation particulière d'un groupement hydroxyle lié à la fonction 17β ol persiste quand on considère une molécule plus volumineuse, comme c'est le cas pour le complexe oestradiol + propanol (orthorhombique $P2_12_12_1$). Bien que l'interaction entre la fonction hydroxyle et le groupement méthyl soit plus faible (3,87$_0$ Å), elle garde la même position dans le "creux" laissé par les hydrogènes du méthyl.





II-3 HYPOTHESES SUR LE ROLE DE LA FONCTION 17β ol

Si les résultats cristallographiques décritsci-dessus peuvent dépendre uniquement de la nature du champ cristallin, (malgré l'existence d'arrangements moléculaires totalement différents), ils pourraient aussi être directement liés au rôle joué par la fonction 17β ol dans l'association du stéroïde avec la protéïne réceptrice.

Deux hypothèses différentes peuvent être formulées à partir des observations cristallographiques :

- on peut supposer que l'eau complète la "structure locale" du stéroïde ; le produit actif fixé sur le récepteur protéïnique étant constitué de l'ensemble des deux molécules (*).

- on pourrait également admettre que le "logement" occupé par la molécule d'eau ou l'oxygène du propanol, reçoit, au moment de la fixation sur le récepteur, une fonction alcool libre de la protéïne (Thréonine ou Sérine) ou encore une fonction thiol (Cystéïne). Comme,d'après les résultats cristallographiques, il semble que le cycle D du stéroïde et l'oxygène alcoolique forment un ensemble "rigide", les liaisons hydrogène supplémentaires contractées par la fonction 17β ol orienteraient alors le substrat dans une position bien déterminée (par le squelette de la protéïne réceptrice) (**).

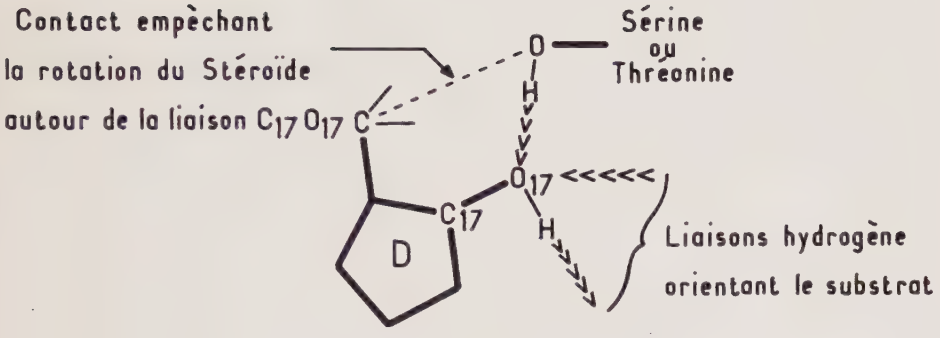
(*) *Dans l'oestradiol hydratée, la distance qui sépare l'oxygène phénolique de l'oestradiol de la molécule d'eau "fixée" sur la fonction 17β ol est égale à 12,24 Å, valeur étrangement proche de la distance $O_{10}\cdots O_{10'}$ du DES (12,13 Å); est-ce réellement fortuit ?*

(**) *Cette dernière hypothèse ne semble pas incompatible avec certains résultats pharmacodynamiques, par exemple :*

- le fait que le 18 nor oestradiol, où l'on a supprimé le groupement méthyl soit 2×10^{-5} fois moins actif que l'oestradiol (R.A. Edgren 1960).

- la diminution notable des constantes d'association avec la protéine réceptrice quand on éloigne la fonction alcool portée par le cycle D stéroïde du groupement méthyl C_{18} (c'est le cas de l'iso oestradiol ou 17α oestradiol ou du 17β CH_2OH 3 hydroxy oestratriène 1, 3, 5). La faible constante d'association observée pour l'oestriol (10 fois plus petite que l'oestradiol) pourrait, dans notre hypothèse, être expliquée par la création d'un pôle d'attraction supplémentaire (16αOH) pour la fonction hydroxyle ou thiol protéinique,plus fort que le méthyl C_{18}, qui amènerait une orientation différente du stéroïde.

- la vitesse d'association nettement plus faible observée pour l'oestrone.



II-4 CONCLUSION

D'autres hypothèses sont sans doute possibles, mais elle devront vraisemblablement faire intervenir l'occupation du "logement" formé par la fonction 17β ol et le groupement méthyl C_{18} par une fonction susceptible de former une liaison hydrogène (-OH, SH ou à la rigueur -NH).

Les hypothèses proposées ne pourront vraisemblablement pas dans tous les cas, rendre compte du mode d'interaction d'un stéroïde avec les protéïnes réceptrices, cependant les associations stéroïde-solvant observées au niveau du cristal présentent une fréquence suffisante pour que l'on puisse espérer retrouver ces schémas d'interactions au niveau de certains récepteurs.

CONCLUSION

L'étude cristallographique d'une série de composés présentant une même activité permet généralement d'avancer certaines hypothèses quant au mode de fixation de ces composés avec leurs protéïnes réceptrices, mais elle apporte trop peu d'éléments pour que l'on puisse conclure avec quelque certitude. Seule la détermination des structures des protéïnes réceptrices peut faire avancer le problème de la fixation des hormones oestrogènes et androgènes de façon décisive. Souhaitons qu'elle permette un jour de répondre aux trop nombreuses questions restant encore en suspens....

BIBLIOGRAPHIE (1ère SECTION)

1 - D.F.HIGH, J.KRAUT - Acta Cryst. (1966) 21 p.88

2 - P.B.BRAUN, J.HORNSTRA, J.I.LEENHOUTS - Philips Research (1969) Vol.24 N°2 p.85

3 - B.BUSETTA - Thèse 3ème Cycle (1968) Bordeaux

4 - B.BUSETTA - Cecam (1970)

6 - B.BUSETTA, M.HOSPITAL - Acta Cryst. B.28 p.560 (1972)

7 - B.BUSETTA, C.COURSEILLE, S.GEOFFRE, M.HOSPITAL - Acta Cryst. B.28 p.1349 (1972)

8 - B.BUSETTA, C.COURSEILLE, M.HOSPITAL - Compte Rendu Acad.Sc.Paris t.271 p.1558 (1970)

9 - B.BUSETTA, C.COURSEILLE, M.HOSPITAL - " " " Sc.Paris t.272 p.1211 (1971)

10 - Résultats à paraître

11 - B.BUSETTA, C.COURSEILLE, J.M.FORNIES-MARQUINA, M.HOSPITAL - Cryst.Struct. Comm. (1972) 1 p.43.

12 - B.BUSETTA, C.COURSEILLE, Mme F.LEROY, M.HOSPITAL - Compt.rend.Acad.Sc. Paris t.272 p. 1304 (1971)

12b- B.BUSETTA, C.COURSEILLE, Mme F.LEROY, M.HOSPITAL - Acta Cryst. B.28 p.3293 (1972)

14 - B.BUSETTA, G.COMBERTON (à paraître)

15 - G.COMBERTON - Diplôme CNAM Bordeaux (1973)

16 - W.COCHRAN- Acta Cryst. (1955) 8 p.473

17 - W.COCHRAN et A.S.DOUGLAS (1955) - Proc.Roy.Soc. A, 227, 486

18 - C.COURSEILLE, B.BUSETTA, M.HOSPITAL - Compt.Rend.Acad.Sc.Paris t.275 p.95 (1972)

19 - A.COOPER, D.A.NORTON, H.HAUPTMAN (1969) - Acta Cryst. B.24 p.814

20 - D.A.NORTON, G.KARTHA, C.T. LU (1964) - Acta Cryst. B.17 p.77

21 - J.FISHER, H.HANCOCK, H.HAUPTMAN - N.R.L. Report (1970)

22 - G.GERMAIN, M.M.WOOLFSON (1968) - Acta Cryst. B.24 p.91

23 - G.GERMAIN, P.MAIN, M.M.WOOLFSON (1970) - Acta Cryst. B.26 p.274

24 - E.GIGLIO (1959) - Nature vol.222 p.339

25 - H.HAUPTMAN, J.KARLE (1956) - Acta Cryst. B.9 p.45

26 - H.HAUPTMAN - The Calculation of Structure Invariants - Parma School of Crystallography (1970)

27 - H.HAUPTMAN, J.FISHER, H.HANCOCK, D.A.NORTON (1969) - Acta Cryst. B.24 p.811

28 - J.L.KARLE et J.KARLE (1969) - Acta Cryst. B.25 p.434

29 - J.KARLE et H.HAUPTMAN (1958) - Acta Cryst.B.11 p.264

29b- J.KARLE et H.HAUPTMAN (1953) - Solution of the Phase Problem I. - The Centrosymetric Crystal ACA Monograph n°3

29c- J.KARLE et H.HAUPTMAN (1957) - Acta Cryst. B.10 p.515

30 - J.KARLE, I.L.KARLE (1966) - Acta Cryst. B.21 p.849

31 - D.SAYRE - Acta Cryst. (1952) 5, p.60

32 - H.SCHENK - Acta Cryst. (1971) B.27 p.2037

33 - P.A.VAUGHAN (1956) - American Crist.Assoc.Annual Meeting French Lick, Ind.

BIBLIOGRAPHIE (2ème Section)

1 - L.S.BARTELL - J.Chem.Phys. 32 p.827 (1960)

2 - E.E. BAULIEU - Nouvelle Revue Fse d'Hématologie Tome 7 (1967) n°5 p.589
Le mécanis.d'action des hormones sexuelles CM 15 III (1969) 91-11 p.2181
Métabolism and Protein binding of sex Stéroïd

3 - G.BROWNLEE and A.F.GREEN - J.Endocrinology 5 - 158 (1947)

4 - R.BUCOURT et D.HAINAUT - Bull.Soc.Chim.France (1965) p.1366

5a- B.BUSETTA - Thèse Pharmacie Bordeaux (à paraître)

5b- B.BUSETTA, C.COURSEILLE, M.HOSPITAL - Acta Cryst. B.29 p.298 (1973)

6 - C.COURSEILLE, B.BUSETTA, G.COMBERTON, M.HOSPITAL
a/ Compte Rendu 272 série C (1971) p.1115
b/ Compte Rendu 274 série C (1972) p.1921

7 - E.C.DODDS, L.GOLBERG, W.LAWSON and R.ROBINSON - Nature 142, p.34 (1938)

8 - S.GEOFFRE (Communication personnelle)

9 - G.GIACOMELLO, E.BIANCHI - Gazz.Chim.Ital. 71 p.667 (1941)

10 - E.GIGLIO - Nature Vol.222 (1969) p.339

11 - J.JACQUES - Bull.Soc.Chim.Fr. 16 p.411 (1949)

12 - B.T.KILBOURN, P.G.OWSTON - J.Chem.Soc. B. p.1 (1970)

13 - A.I.KITAIGORODSKI - Advances in Structure Research by diffraction methods vol.3 p.173

14 - H.P. KOCH - Nature 161 p.309 (1948)

15 - K.S. PITZER - Discuss.Faraday Soc. 1951 n°10 p.66

16 - B.et A.PULLMAN - Les Théories Electroniques de la Chimie Organique (1952)

17 - L.A. TCHETKINA et G.A.GOLDER ЖУРН. СТРУКТ. ХИМ - Tom 9, n°2 p.254 (1968)

18 - TOMKINS

19 - UNANUE - Thèse d'Etat Bordeaux (1965)

20 - UGNADE et MORISS - J.Amer.Chem.Soc. (1947)

21a- C.M.WEEKS, A.COOPER, D.A.NORTON - Acta Cryst. B.26, 429 (1970)

21b- C.M.WEEKS, J.FISHER, A.COOPER, D.A.NORTON, H.HAUPTMAN - Acta Cryst.B.27, 1562 (1971)

22 - G.ANNER, J.KALDOVA, T.WIELAND - Ciba Corp. New York - U.S.Patent 3,318, 927 (1967)

23 - D.BOURQUIN, G.AZADIAN BOULANGER, D.PHILIBERT, J.P.RAYNAUD - Thild Int Congr.Hormonal Stéroïd 210 p.312 (1970)

24 - G.PRECIGOUX - Thèse 3° Cycle Bordeaux (1973)

25 - W.C.HAMILTON, J.A.IBERS (1968) Hydrogen Bonding in Solids New York Benjamin

Vu et approuvé
Bordeaux, 26 Mai 1973
pour le Président du Conseil de l'Université de Bordeaux I
le premier Vice-Président
M.RIMPAULT